A Survey of Data Mining Techniques for Analyzing Crime Patterns
Ilamvaluthi Joseph. S1, Bala Subramani. C2, Harish. T3, Mohan raj4

Student1,2,3, Assistant professor4
Department of Computer Science and Engineering
Velammal Engineering College, Chennai, India

Abstract:
In recent years the data mining is data analyzing techniques that used to analyze crime data previously stored from various sources to find patterns and trends in crimes. In additional, it can be applied to increase efficiency in solving the crimes faster and also can be applied to automatically notify the crimes. However, there are many data mining techniques. In order to increase efficiency of crime detection, it is necessary to select the data mining techniques suitably. This paper reviews the literatures on various data mining applications, especially applications that applied to solve the crimes. Survey also throws light on research gaps and challenges of crime data mining. In additional to that, this paper provides insight about the data mining for finding the patterns and trends in crime to be used appropriately and to be a help for beginners in the research of crime data mining.

Index Terms: Data mining, Crime analysis, Big data.

5721 . INTRODUCTION

Big data is the voluminous and complex collection of data that comes from different sources such as sensors, content posted on social media website, sale purchase transaction etc. Such voluminous data becomes tough to process using ancient processing application. There are various tools and techniques in the market for big data analytics. With continually increasing population, crimes and crime rate analyzing related data is a huge issue for governments to make strategic decisions so as to maintain law and order. This is really necessary to keep the citizens of the country safe from crimes. The best place to look up to find room for improvement is the voluminous raw data that is generated on a regular basis from various sources by applying Big Data Analytics (BDA) which helps to analyze certain trends that must be discovered, so that law and order can be maintained properly and there is a sense of safety and well-being among the citizens of the country. 90% of the world’s data was generated in the last few years. Due to the advent of new technologies, devices, and communication means like social networking sites, the amount of data produced by mankind is growing rapidly every year. The amount of data produced by us from the beginning of time till 2003 was 5 billion gigabytes. If you pile up the data in the form of disks it may fill an entire football field. The same amount was created in every two days in 2011, and in every ten minutes in 2013. This rate is still growing enormously. Though all this information produced is meaningful and can be useful when processed, it is being neglected. Big Data is a collection of large datasets that cannot be processed using traditional computing techniques. It is not a single technique or a tool rather it involves many areas of business and technology. This software provides facility for reporting online crime, complaints, missing persons, show most wanted person details mailing as well as chatting. Any Number of clients can connect to the server. Each user first makes their login to sewer to show their availability. The server can be any Web Server. An SMTP Server must be maintained for temporary storage of emails and chat jar files for enable the chatting facilities. The main modules in this project. Big data is the voluminous and complex collection of data that comes from different sources such as sensors, content posted on social media website, sale purchase transaction etc. Such voluminous data becomes tough to process using ancient processing application. There are various tools and techniques in the market for big data analytics. With continually increasing population, crimes and crime rate analyzing related data is a huge issue for governments to make strategic decisions so as to maintain law and order. This is really necessary to keep the citizens of the country safe from crimes. The best place to look up to find room for improvement is the voluminous raw data that is generated on a regular basis from various sources by applying Big Data Analytics (BDA) which helps to analyze certain trends that must be discovered, so that law and order can be maintained properly and there is a sense of safety and well-being among the citizens of the country. In this paper we are analyzing crime data in different areas in a city by using Hadoop tool along with some Hadoop ecosystems like HDFS, Map Reduce, SQOOP, hive and pig. By using these tools we can process no limitation of data, no data lost problem, we can get high throughput, maintenance cost also very less and its a opensource software, it is compatible on all the platforms since it is Java based.

II. EXISTING SYSTEM

Existing concept deals with providing backend by using ysql which contains lot of drawbacks i.e data limitation is that processing time is high when the data is huge and once data is lost we cannot recover so thus we proposing concept by using Hadoop framework. It is important to study the previous related works to both learn from the experience of others and to add something to our existing body of knowledge. Existing literature has been reviewed in three different areas: crime data mining, data extraction and data focus. Crime Data Mining: Data mining is defined as the discovery of interesting structure in data, where
structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data. The data mining techniques is using for some results on crime mining. This technique is applied to study crime cases, which mainly concerned entity extraction, pattern clustering, classification and social network analysis. This method used to get the data of criminals by using frequency occurrence of incidents. **Data Extraction Event**: In web page extraction is the process to extract attributes and relationship. The idea of this event extraction is the method of retrieving the information from database. The proposed a method to append events for the concept of data mining techniques. Entity extraction has been used to automatically identify person, address, cases, and personal properties from police reports to the judge. **Data Focus**: In this research paper is on web mining of content, using clustering techniques the web mining focus on the text. In data focus clustering will convert nonlinear statistical relationship between high dimensional data into simple geometrical relationship in low dimensional display.

1.2 DISADVANTAGE
- We can process limitation of data.
- We get results with take more time and maintenance cost is very high.

III. PROPOSED SYSTEM

Proposed concept deals with providing database by using Hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster and we get results with less time, high throughput and maintenance cost is very less and we are using joins, partitions and bucketing techniques in Hadoop. Hadoop is an open source framework which also by the apache software foundation and it is used for storing and processing huge datasets with a cluster of commodity hardware. We use Hadoop tool contains two things one is HDFS and MapReduce. We also use Hadoop ecosystems like SQOOP, Hive and Pig.

2.2 ADVANTAGE
- No data loss problem
- Efficient data processing.

IV. SYSTEM ARCHITECTURE

![System Architecture](image)

Figure.1.System Architecture

V. MODULE DESCRIPTION

Modules:
- Data Preprocessing Module
- Data Migration Module With SQOOP
- Data Analytic Module With Hive
- Data Analytic Module With Pig
- Data Analytic Module With Map Reduce

Data Preprocessing Module

In this module we have to create Data set for Feedback dataset it contain set of table such that User details, Baby Item details, Electronic Item details, sports Item details, Men Item Details and feedback or transactions details for last one years and this data first provide in MySQL database with help of this dataset we analysis this project.

Data Migration Module with Sqoop

Now we are ready with dataset. So now our aim is transfer the dataset into Hadoop (HDFS), that will be happen in this module. Sqoop is a command-line interface application for transferring data between relational databases and Hadoop. In this module we fetch the dataset into hadoop (HDFS) using sqoop Tool. Using sqoop we have to perform lot of the function, such that if we want to fetch the particular column or if we want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in hadoop (HDFS).

Data Analytic Module with Hive

Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions. In this module we have to analysis the dataset using HIVE tool which will be stored in hadoop (HDFS). For analysis dataset HIVE using HQL Language. Using hive we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language.

Data Analytic Module with Pig

Apache Pig is a high level data flow platform for execution Map Reduce programs of Hadoop. The language for Pig is pig Latin. Pig handles both structure and unstructured language. It is also top of the map reduce process running background. In this module also used for analyzing the Data set through Pig using Latin Script data flow language. in this also we are doing all operators, functions and joins applying on the data see the result.

Data Analytic Module with MapReduce

MapReduce is a processing technique and a program model for distributed computing based on java. The Map Reduce algorithm contains two important tasks, namely Map and Reduce. In this module also used for analyzing the data set through MAP REDUCE. Map Reduce Run by Java Program.

VI. HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a
complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the system does and not how it should be implemented.

**SOFTWARE REQUIREMENTS**

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

<table>
<thead>
<tr>
<th>Framework</th>
<th>Hadoop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>MY SQL 5.5</td>
</tr>
<tr>
<td>Language</td>
<td>Pig, Hive, CoreJava</td>
</tr>
<tr>
<td>Data Access Tool</td>
<td>Sqoop</td>
</tr>
<tr>
<td>Operating System</td>
<td>cent os</td>
</tr>
<tr>
<td>IDE</td>
<td>Eclipse</td>
</tr>
</tbody>
</table>

**VIII. CONCLUSION**

Big Data Analytics refers to the tools and practices that can be used for transforming this raw data into meaningful and crucial information which helps in forming a decision support system for the judiciary and legislature to take steps towards keeping crimes in check. With the ever increasing population and crime rates, certain trends must be discovered, studied and discussed to take well informed decisions so that law and order can be maintained properly. If the number of complaints from a particular state is found to be very high, extra security must be provided to the residents there by increasing police presence, quick redressed of complaints and strict vigilance. Crimes against women are becoming an increasingly worrying and disturbing problem for the government.

**IX. REFERENCE**


