CEKRIT: Cyber Enforcement using Knowledge Redundancy and Information Technology

[1]Chinmay Nagvekar , [2]Kalpesh Surve, [3]Madhuri Mistry, [4]Varsha Wangikar
1- Student 2-Student 3-Student 4-Assistant Professor,
K.C.College of Engineering and Management Studies and Research,
Mith Bunder road, Near Sadguru Garden, Kopri, Thane East- 400603.
India.

Abstract : Crimes are increasing now day by day and cyber crimes are as such which takes a lot of time to be processed and give solutions . This research work attempts to minimize the burden and tedious process of registering and giving solution for similar cyber crimes to certain extent.

***Keywords*—Prediction, Gradient Descent , Stemming , NLP , Cyber Crimes**

I. INTRODUCTION

Cyber crime is one of the fastest growing crimes of 2018. It mostly involves all kind of computer and network related crimes. This system provides an easy way to manage and handle cyber crime cases. This project aims at creating a system which can help an investigating officer to access and go through previous records and cases which are similar to the current one and give solutions quicker. Here we are using Stemming Algorithm for NLP and Gradient Descent algorithm mainly. Using the proposed system the authorized officer can see the previous cases and the areas(location) where the crimes have occurred and give solutions to it faster. This project has a good impact on the efficient way of working of the cyber crime department in the manner of registering a case to solving it completely and helping the victim.. The idea was inspired by reviewing a few IEEE papers based on the crime pattern detection using historical data and geographical crime rates which were focused on the areas of where the crimes were occurring. The main aim was to create an intelligent system which can give solutions by itself after training the initial model with a particular amount of data.

The Hierarchical algorithm is the clustering algorithm used in this project. It is similar to K-means but in an advanced format for the data mining. The Gradient Descent algorithm is used for the major two modules of the project i.e. Decision Making and Prediction of what type of crime can occur where along with its solutions.

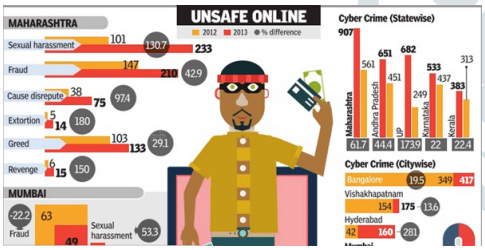

Fig. 1.1 Cyber Crime Rates in Recent time

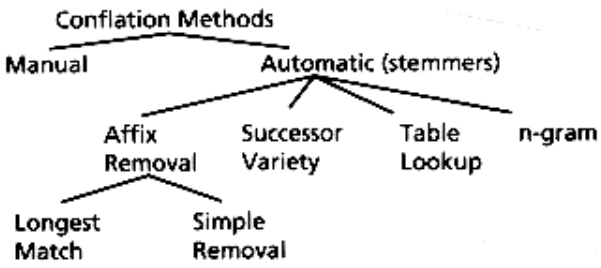

Fig. 1.2 Stemming algorithm for NLP(Natural Language Processing)

A stemming algorithm is a process of linguistic normalization, in which the variant forms of a word are reduced to a common form. It is important to realize that the stemming process cannot be made perfect. Linguistic irregularities slip through the net of a stemming algorithm.

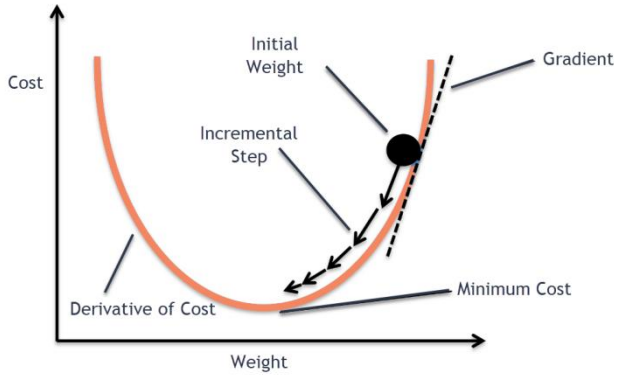

Fig. 1.3 Gradient Descent Algorithm

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point. If, instead, one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent. Gradient descent is also known as steepest descent.

II. LITERATURE SURVEY

A. Bridging the gap between predictive and prescriptive analytics - new optimization methodology needed Dick den Hertog Department of Econometrics and Operations Research, Tilburg University, d.denhertog@tilburguniversity.edu Krzysztof Postek Faculty of Industrial Engineering and Management, Technion, Israel Institute of Technology, krzysztofp@technion.ac.il[1]

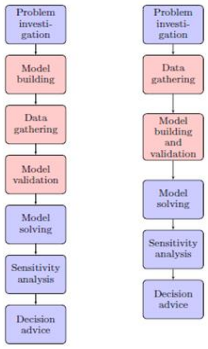

Fig. 2.1 Comparison of the classical model building(left) and the new approach(right)
Business analytics is becoming more and more important nowadays. Up to now predictive analytics appears to be much more applied in practice than prescriptive analytics. We argue that although optimization is used to obtain predictive models, and predictive tools are used to forecast parameters in optimization models, still the deep relation between the predictive and prescriptive analytics is neither well understood nor fully exploited. We describe two opportunities to really exploit the synergy between the predictive and prescriptive part. The rest is to perform optimization by directly using the predictive models. Adding optimization functionality in predictive analytics tools could be of huge added value for practice. The second opportunity is to replace manual model building with automated data-driven model building, using modern predictive analytics. The pros and cons for such a way of optimization are also discussed.

B. Crime pattern detection using historical Data (www.leadingindia.ai)[2]

The crime rate prediction strategies can be applied on historical data available in the police records by examining the data at various angles like reason of crime, frequency of similar kind of crimes at specific location with other parameters to prepare model the crime prediction. It is the major challenge to understand the versatile data available with us then model it to predict the future incidence with acceptable accuracy and further to reduce the crime rate.

Fig. 2.2 Crimes on historical data

This representation is outcome of analysis on the existing crime records. These crimes are majorly related to robbery, murder, burglary which is having higher probability where proper security is not implemented.

This type of analysis on historical data can also help in determining the criminal profile based on the characteristic behavior inferred from the data. It can help the investigator to accurately predict the profile of unknown criminals.

C. Geographical crime rate prediction (www.leadingindia.ai)[3]

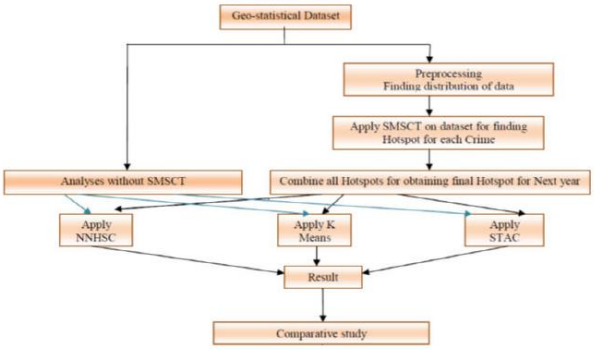

Fig. 2.3 General Framework for crime prediction model

Dataset considered for the proposed solution is the data set of metropolitan cities of India for minimum four years. Data is stored in form of longitude and latitude information of crime location and frequency of such incidences regarding different types of crimes. Data preprocessing is performed to extract and understand the distribution of data. For the purpose of preprocessing convex hull with standard deviation cab be deployed. Convex hull gives boundary wall around the points of distribution and deviation may describe the nature of data from mean point of view.

III. PROPOSED SYSTEM

Based on the above literature survey we have decided our solution as given below.

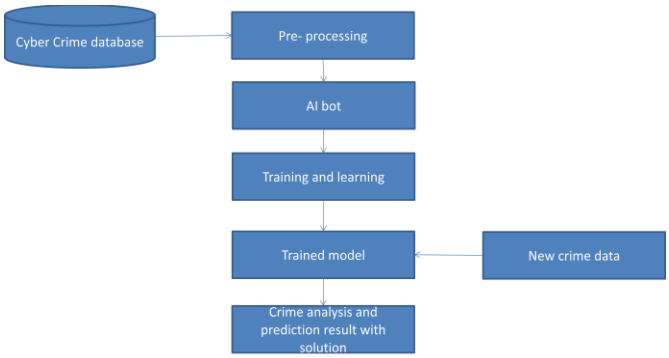

Fig. 3.1 Basic working (Block Diagram)

The above is the basic working of the flow of the project. Huge amount of database is connected to an AI based bot which is then trained with the available data and solutions of already occurred crimes. This trained model will be updated with new crime data time to time. Crime analysis and prediction will be done along with solution by the AI bot.

IV. METHODOLOGY

The main aim is to create a system that will cluster the cyber crime database according to the types and areas. The next step is to give this clustered data to the solution mechanism that uses advanced version of the Hierarchical Algorithm and decision tree algorithms to produce efficient solutions for the cyber crimes. The final module solely comprises of the Gradient Descent algorithm to predict further cyber crime on the basis of the cyber crime types and the area in which they are committed.

Fig. 4.1 Natural Language Processing

Crime can be better predicted if we use clustering approach rather than classification due random nature of crimes. Clustering approach will work better. There is one algorithm which we have proposed after trying several others is the Stemming algorithm for NLP(Natural language processing) . Stemming Algorithm has better accuracy and results instead of K-means and logistics regression.

The method we are using directly is Gradient Descent for Predictive Analysis.

Following steps are applied by us for the crime analysis:

1. Collect the data:

In this phase the data from various sources is collected from various government sources, social media platform about the incidence like facebook, blogs etc. It may expect that the data received may be in unstructured form with different types and size. So oops approach may be better to extract the information explicitly.

2. Classification:

The direct methodology used for classification are specifying the keywords and the categories of the crimes.

3. Identify pattern:

Next phase in the methodology is to find the sequence of crimes which are similar in

nature and belongs to same class. Such pattern may be identified as suggested in the Stemming algorithm for natural language processing.

Pattern identification outcome may be location based pattern detection. Pattern

information may help the government or police department to effectively speedily

work on resolving the criminal cases.

4. Prediction of Crime:

For the prediction outcome Gradient Descent is used. Here in this type of prediction the results are to be estimated only if the unit comes 0.5 and above then the results will be displayed in the form of a graph. If the unit comes below or less than 0.5 then the graph will be negligible in that case.

5. Visualization or presentation of result:

The results can be visualized using appropriate

graphs or maps showing sensitive areas of having high probability of crimes.

V. IMPLEMENTATION

When the main code will be executed, case description will be taken as an input. On this input we will apply the stemming algorithm. It will result into a tree form on which traversal has to be performed. This traversal has to be performed on certain keywords.

data.json

```
{
    "1": ["debit card","credit card","atm","card"],
    "2": ["bank","otp"],
    "3": ["insurance"],
    "4": ["paytm","wallet","payumoney","mobikwik"]
}
```

Fig. 5.1 CEKRIT Keywords

The above given image 5.1 shows the keywords on which that traversal has to be performed. The traversal will be done on the basis of keywords and then output binding has to be done.

Training simulations has to be given to the AI bot in the form of the training sets. A minimum of 1000 epoch simulations has to be done.

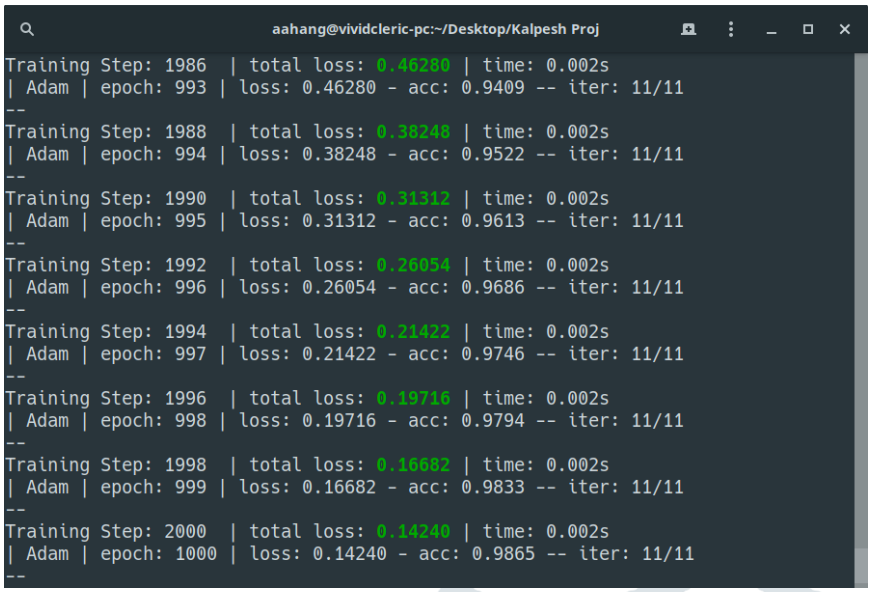

Fig. 5.2 Training Simulations

After getting these simulations the prediction and solution part of the project comes in.

The next image shows how the result or final output will be displayed in the system.

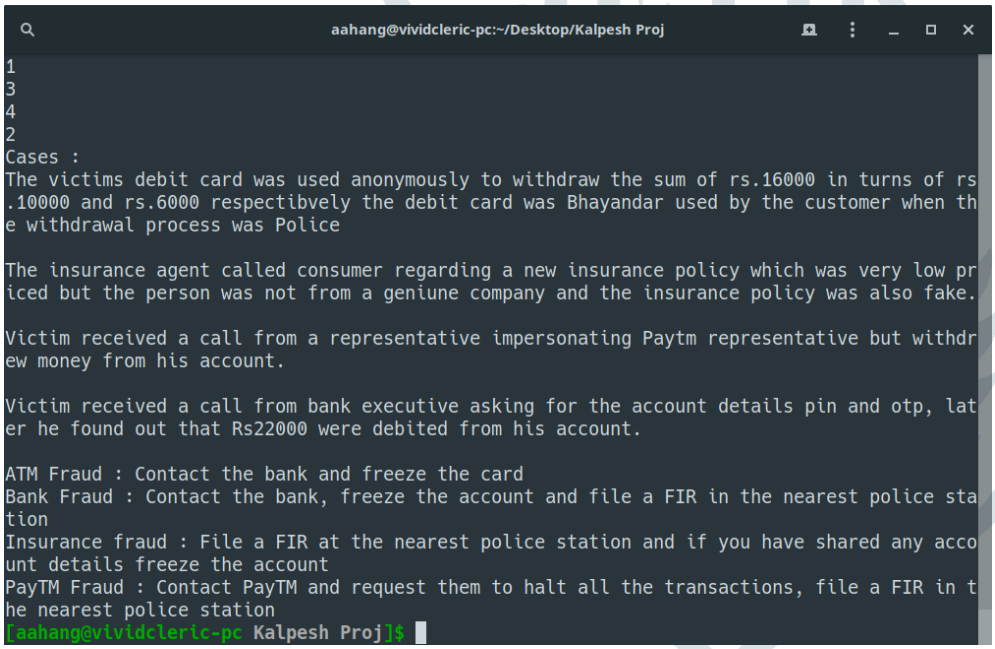

Fig. 5.3 Output Screen for CEKRIT

The output shows case description and the possible solutions to each cases.

VI. RESULTS ANALYSIS

From the above given Fig. 5.1, 5.2 and 5.3 we can see how the analysis of one case is done and solution is given by the system using the keywords and the simulation process. The output shows the case description and the solution for the given case and also gives out the prediction.

VII. CONCLUSION

In conclusion we would like to state that this system is a Cyber Crime detection and solution providing system which can give solutions to each case on its own once it is trained and fed with the previous databases. It is an optimal solution to the Cyber Cell branch of our state. We are developing this project on the above given criteria and guidelines. It can further be updated with a lot more things. Predictive analysis may be improved in future with added features of the new technology.

VIII. ACKNOWLEDGEMENT

We express sincere thanks to our guide Prof. Varsha Wangikar whose supervision, inspiration and valuable guidance helped us a lot to complete our work. Her guidance proved to be the most valuable to overcome all the hurdles in the fulfillment of this paperwork. Also, we are thankful to all those who have helped us in the completion of paperwork.

IX. REFERENCES

[1] "Bridging the gap between predictive and prescriptive analytics - new optimization methodology needed Dick den Hertog Department of Econometrics and Operations Research, Tilburg University, d.denhertog@tilburguniversity.edu Krzysztof Postek Faculty of Industrial Engineering and Management, Technion, Israel Institute of Technology, krzysztofp@technion.ac.il "

[2]"Crime pattern detection using historical Data"(www.leadingindia.ai)

[3]"Geographical crime rate prediction"(www.leadingindia.ai)