

Predictive Policing: Preventing Crime with Data Analytics

Aditya Sharma¹, Vivek Kumar Prasad²

¹(Information Technology, Institute Of Technology, Nirma University, Ahmedabad, India

²(Computer Science and Engineering Department, Institute Of Technology, Nirma University, Ahmedabad India

ABSTRACT

This paper is an attempt to study the idea of using predictive models and visualisation to prevent crime. It involves the use of different predictive models applied on previous data (reports and cases of crime) in determining the probable crime at a particular location, at a particular day of the week and the time of crime. By using visualisation, police officers can easily analyse the location where the probability of crime is high and can prevent it from happening. Visualisation acts as a medium between predictive model and police officers.

Keywords: Data Analytics, Forecasting, Hot Spot Clustering, Predictive model, Visualisation.

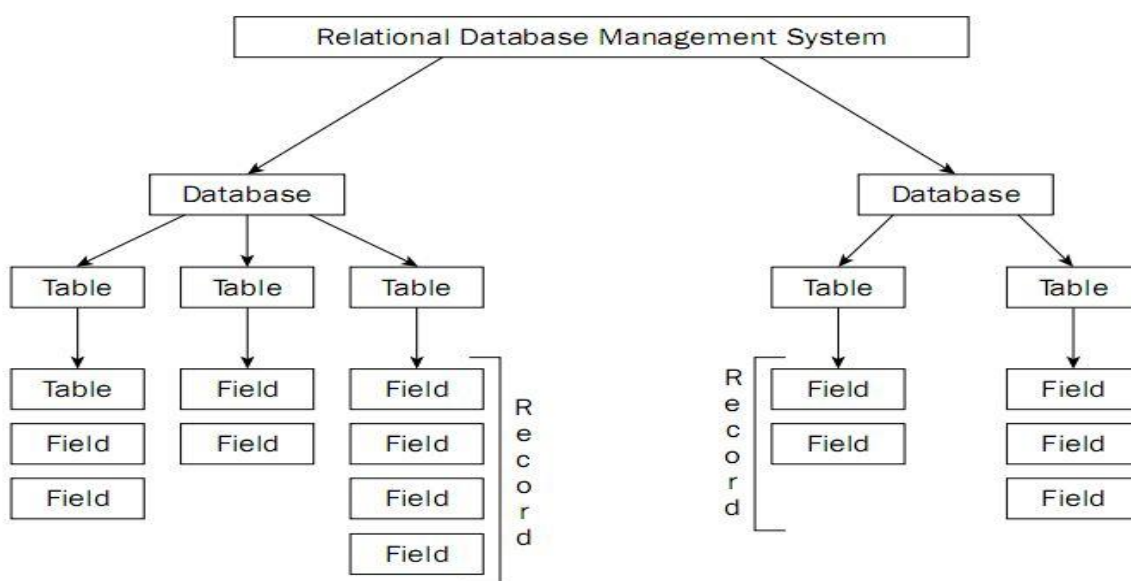
I. INTRODUCTION

To apply predictive policing we need the data to be well maintained and digitalized before applying any model on it. We need RDBMS (relational database management system) to store data and collaborate data from different geographical locations, so that we can apply Data Warehousing techniques on the given data set to extract relevant data on which we can apply different predictive models. Relevant data can include the type of crime (burglary, theft, property crime, violent crime, auto-theft, extortion, etc), date of crime, time at which crime happened and location of crime scene.

Using this data, we can apply predictive models and the result obtained from the model can be used to forecast the location, type and time of crime and map it onto Google Maps so that police officers can understand the forecast without having the knowledge of the process applied in forecasting the crime.

II. DATABASE MANAGEMENT

The primary task in applying predictive policing is database management. We need to digitalize our data and abandon the traditional method of record keeping so that the data is easily available whenever required and operations can be performed on the data set. By using RDBMS (relational database management system), database can be partitioned on the basis of different geographical regions. Each region can use the data relevant to that region by accessing the database. By using RDBMS (relational database management system), we achieve consistency in data which would help us in making a better model for prediction.



After database management, we require to perform different Data Warehousing techniques so that our data is in relevance to the task to be performed. Applying a predictive model after applying Data Warehousing on data gives a better prediction score than that performed on raw data. Time required by predictive model is also reduced by using Data Warehousing, resulting in better efficiency.

III. PREDICTIVE MODEL

Different predictive models are applicable for different types of data. Some data sets may require normalization. Depending on the requirement, some of the predictive models are:-

i. ONE VARIABLE LINEAR REGRESSION MODEL

$$: Y^i = \beta_0 + \beta_1 X^i + \gamma^i \quad (1)$$

Y^i = dependent variable for the i^{th} observation

X^i = independent variable for the i^{th} observation

γ^i = error term for the i^{th} observation

β_0 = intercept coefficient

β_1 = regression coefficient for the independent variable

ii. MULTIPLE VARIABLE LINEAR REGRESSION MODEL

$$: Y^i = \beta_0 + \beta_1 X_1^i + \beta_2 X_2^i + \dots + \beta_k X_k^i + \gamma^i \quad (2)$$

Y^i = dependent variable for the i^{th} observation

X_j^i = j^{th} independent variable for the i^{th} observation

γ^i = error term for the i^{th} observation

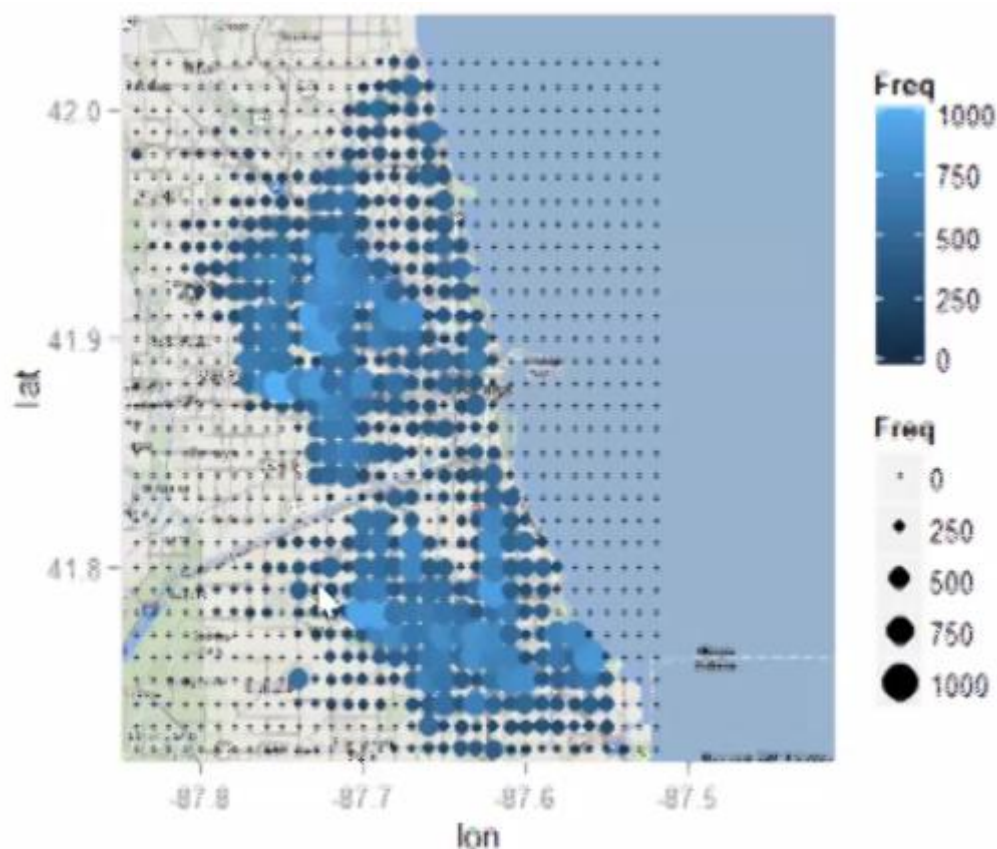
β_0 = intercept coefficient

β_j = regression coefficient for the j^{th} independent variable

Different predictive models like CART, randomFOREST, etc. are also applicable. Clustering technique is used with these predictive models to attain a better predictive score.

IV. VISUALISATION

The data points obtaining a predictive score greater than predefined value is given a value TRUE. The coordinates of the data points holding value TRUE are mapped onto Google Maps. By applying different visualisations like heat cluster map we can analyse which locations on the map are most affected by crimes and preventive measures can be applied to lower the crimes in those locations. By using “ggmap” package in R we can extract the map of a particular region (city, state, country or continents) and map the coordinates on the extracted part for a better visualisation analysis.



V. CONCLUSION

- By using RDBMS we can easily manage the data and access it.
- Data Warehousing reduces the inconsistency in data providing a better data set for further operations.
- By using proper predictive models we can reach the best possible predictive score.
- By merging Predictive modelling and Visualisation we can attain the best possible Predictive Policing Model. This would result in prevention of crimes in the best possible manner.
- This model can also be give vital information to people who want to move to a location where number of crimes are low.
- It would result in better utilisation of resources available to the police force. Cops can reach the scene of crime even before the crime is committed. Thus, reducing the crimes in that location.
- Less crimes would mean less number of cases, which would ease off the load from the Judiciary System. This would result in lesser pending cases and people could get justice faster.

REFERENCES

Image:

[1] <http://img.c4learn.com/2012/02/Relational-Database-Management-System.jpg>

Course:

[2] <https://www.edx.org/course/analytics-edge-mitx-15-071x-0>