Data Analysis of Crime using Machine

Learning Algorithms

**Fakhruddin Tahery , B M Tanvi Raj , Arshi Khan , B.Nishanth Varma**

Students ,Computer Science ,MSRIT , Bangalore.

**Abstract**: Data mining and machine learning play a vital role in early crime detection and therefore, its subsequent prevention. We implement the regression and polynomial regression algorithms using the same finite set of features, on the crime dataset. The main objective of the project is to provide time-based, location-based and crime-type based rate and count ,which may help the society to avoid the highly concentrated places of crime.

Keywords: Polynomial Regression, Subsequent Prevention.

**Introduction**

The Crime reporting program categorizes the offenses into four categories: murder, forcible rape, robbery, and aggravated assault. Murder basically is the wilful killing of one human being by another. Forcible rape is a sexual attack on a female against her will. Though attempts or assaults to commit rape by threat or force are considered crime under this category, statutory and other sex offenses are excluded. Robbery is the taking or attempting to take anything of value from the care, custody, or control of a person or persons by force or threat of force or violence and/or by putting the victim in fear. Aggravated Assault - It is the unlawful attack conducted by one person upon another to inflict severe or aggravated bodily injury.

In this research we implemented polynomial aggression to predict the crimes with respect to the type, count and rate and also using SAP Lumira for the plotting of graphs to get the overall review.

The objectives of this implementation of idea are to Predict and provide preventive measures for the citizens in order to avoid the possible crime situations, present a list to the society the types of crime possible at particular places and also to give a chance to an individual to avoid certain heat mapped places of particular crimes and to provide type, rate and count of any crime in chronological order.

The security issues are covered by the system’s back-end servers shall only be accessible to authenticated administrators. It provides safeguards to prevent damage to data from operator errors, simultaneous

Updates , module unavailability or system failures etc.

The application is portable. The software can be used on different environments .It is adaptable to different specified environments without applying actions or means other than those provided for this purpose for the system. The application is portable across multiple operating systems and Browsers.

It is also reliable as Application will not fail without manual intervention. The application will be error-free and will be available all the time if no manual intervention took place.

**ALGORITHM USED**

**Polynomial Regression**

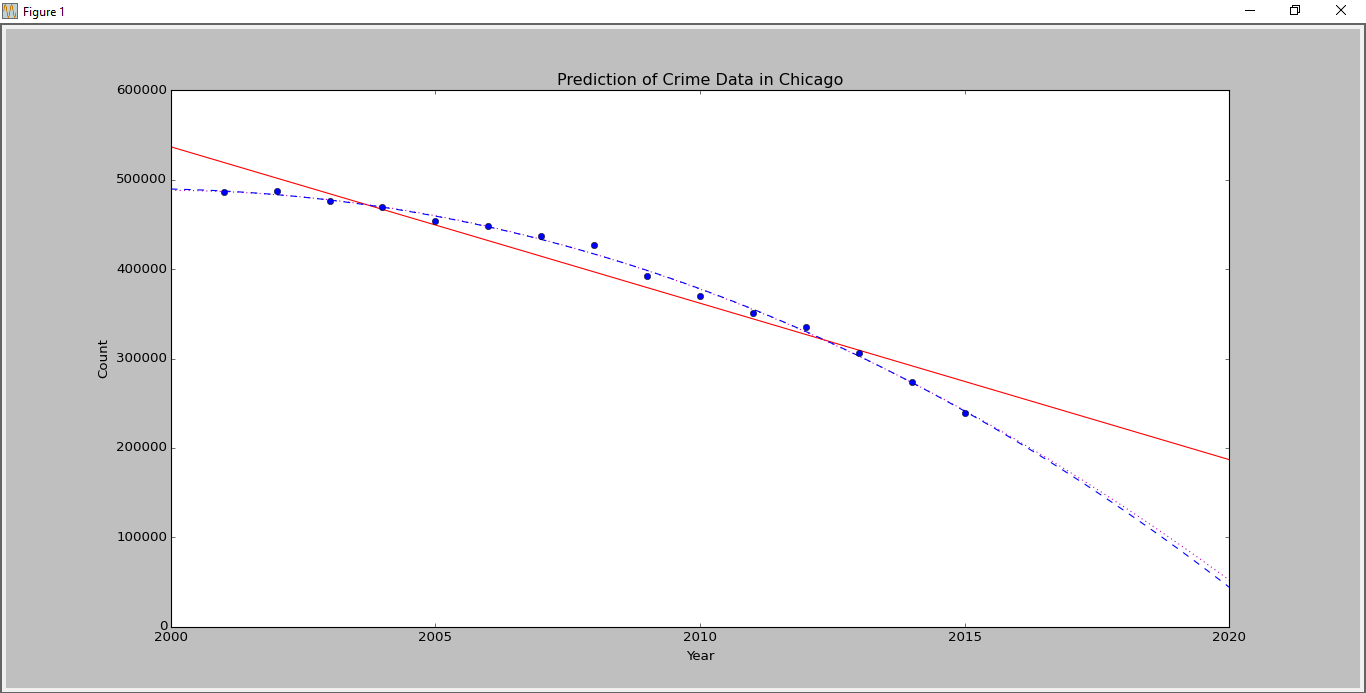
It is a form of [linear regression](https://en.wikipedia.org/wiki/Linear_regression) in which the relationship between the independent variable *x* and the dependent variable *y* is modelled as an *n*th degree [polynomial](https://en.wikipedia.org/wiki/Polynomial). Polynomial regression fits a nonlinear relationship between the value of *x* and the corresponding [conditional mean](https://en.wikipedia.org/wiki/Conditional_expectation) of *y*, denoted E(*y* | *x*), and has been used to describe nonlinear consequences. We have used this algorithm to predict the crime type, rate and count for any particular location in Chicago with longitude and latitude.

**SAP Lumira**

SAP Lumira (formerly called SAP Visual Intelligence) is a self-service, data visualization application .Here in this analysis we use this application for plotting a graph for consecutive years over a period of time.

We use this particular tool to plot the graph which gives a picture how the prediction has a chance to take place.

Fig. 1. Prediction



**DATASETS**

The data sets retrieved for the further implementation is from “data.cityofchicago.org”.

Columns are present describing the following areas

of interest :Year ,Primary type ,Location ,Longitude

Latitude ,Arrest [ Yes or No ]

**CONCLUSION**

We observe the polynomial regression algorithm to be very effective and accurate in predicting the crime data based on the training set input for the algorithm. Utilizing these applications of data mining can be a long and tedious process for law enforcement officials who have to sift through large volumes of data. However, the precision in which one could infer and create new knowledge on how to slow down crime is well worth the safety and security of people .There are other applications of data mining in the realm of law enforcement such as determining criminal "hot spots", creating criminal profiles, and learning crime trends.

**ACKNOWLEDGMENT**

We would like to thank our institution M. S. RAMAIAH

INSTITUTE OF TECHNOLOGY for giving us this opportunity to do a project on Data Analytics in the seventh semester.

**REFERENCES**

[1]ViolentCrime.http://www2.fbi.gov/ucr/cius2009/offenses/violent\_crime/.

[2]Murder.http://www2.fbi.gov/ucr/cius2009/offenses/violent\_crime/murder\_homicide.html.

[3]ForcibleRape. <http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/forcible_rape.html>.

[4]Robbery.http://www2.fbi.gov/ucr/cius2009/offenses/violent\_crime/robbery.htm

[5]Assault.http://www2.fbi.gov/ucr/cius2009/offenses/violent\_crime/aggravated\_assault.html.

[6]Mississippi Crime Rates and Statistics - NeighborhoodScout. Mississippi Crime Rates and Statistics - NeighborhoodScout. Accessed February 17, 2015. <http://www.neighborhoodscout.com/ms/crime/>.

[7] S. M. Nirkhi, R.V. Dharaskar and V.M. Thakre. "Data Mining : A Prospective Approach for Digital Forensics," International Journal of Data Mining & Knowledge Management Process, vol. 2, no. 6 pp. 41-48, 2012.

[8] Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/.

Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015 12

[9] E. W. T. Ngai, L. Xiu, and D. C. K. Chau. "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review andClassification," Expert Systems with Applications, pp. 2592–2602, 2008.

[10] J. McCarthy, "Arthur Samuel:Pioneer in Machine Learning," AI Magazine, vol. 11, no. 3, pp. 10-11, 1990. [11] R. Bermudez, B. Gerardo, J. Manalang and B. Tanguilig, III. "Predicting Faculty Performance Using Regression Model in Data Mining," Proceedings of the 9th International Conference on Software Engineering Research, Management and Applications, pp. 68-72, 2011. [12] S. Sathyadevan and S. Gangadharan. "Crime Analysis and Prediction Using Data Mining," Proceedings of the 1st International Conference on Networks and Soft Computing, pp. 406-412, 2014. [13] Communities and Crime Un normalized Dataset. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Communities and Crime Unnormalized.