**Literature review**

Understanding how to control crime and make a crime prediction are critical issues around the world because exposure to violence and crime has substantially increased due to the well-developed technology. And social media became a critical tool to prevent crime and make crime prediction, social media allows people to interact with each other, with express their behavior and feeling.

The number of sources for such data is increasing with the more widely used being Twitter, mobile device data from service providers, public transport usage, Foursquare, Flickr, and Facebook. Research in the United States has found that that two-thirds of online adults (66 percent) use social media platforms (Smith, 2011) and that 26 percent of American Internet users aged 18-29 have been found to use Twitter (Smith and Brenner, 2012)

With using the messages from mobile devices (such as smart phones) that are posted to Twitter. These data have the potential to represent the ambient population at much higher spatial and temporal resolutions than previous research in spatial crime analysis. (Malleson, N., & Andresen, M. A. ,2014)

Twitter corpora is a major research area and provide direction for work. Many types of research have been carried out in this aspect. Many research papers examine the using tweet to perform sentiment analysis, for example, stock price trend, human mobility.

These services allow users to instantly create, disseminate, and consume information from any location with access to the Internet. (Wang, X, Gerber, M. S, & Brown, D. E , 2012)

Most sheriffs’ office and police departments use electronic systems for crime reporting that have replaced the traditional paper-based crime reports. These crime reports have the following kinds of information categories namely - type of crime, date/time, location etc (Nath, S. V. 2006). With using the suspect data, we can use the characteristic or feature to predict the crime pattern to prevent next crime happen.

Cluster (of crime) has a special meaning and refers to a geographical group of crime, i.e. a lot of crimes in a given geographical region. (Nath, S. V. 2006).

The aim here is that we can use data mining to detect much more complex patterns since in real life there are many attributes or factors for crime and often there is partial information available about the crime. In a general case it will not be easy for a computer data analyst or detective to identify these patterns by simple querying. Thus clustering technique using data mining comes in handy to deal with enormous amounts of data and dealing with noisy or missing data about the crime incidents (Nath, S. V. ,2006)

Classification algorithms that are mostly used in predictions basing on historical data. Classification is a class prediction technique, which is supervised in nature. (Ahishakiye, E., & Taremwa, D. ,2017)

Clustering is the process of class discovery, where the objects are grouped into clusters and the classes are unknown beforehand. Two clustering techniques, K-means and DBScan (Density-Based Spatial Clustering Application with Noise) algorithm are considered for this purpose. (A. Malathi, Dr. S. Santhosh Baboo ,2011)

By gathering the geographical data, we can use a different method to predict the crime rate or crime pattern, for instance, K mean clustering, Support Vector Machine, or even Random Forest Classifier to classifier the crime pattern

**Reference**

Ahishakiye, E., & Taremwa, D. (2017). Crime Prediction Using Decision Tree (J48) Classification Algorithm. http://idr.kab.ac.ug/xmlui/bitstream/handle/20.500.12493/113/Crime%20Prediction%20Using%20Decision%20Tree%20%28J48%29%20Classification%20Algorithm.pdf?sequence=1&isAllowed=y

A. Malathi, Dr. S. Santhosh Baboo (2011). Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.301.9760&rep=rep1&type=pdf>

Smith, A. (2011). Why Americans use social media. Technical report, Pew Research Centre Available online: http://www.pewinternet.org/Reports/2011/Why-Americans-UseSocial-Media.aspx

Smith, A., and J. Brenner. (2012). Twitter Use 2012. Technical report, Pew Research Center Available online: <http://pewinternet.org/Reports/2012/Twitter-Use-2012.aspx>

Malleson, N., & Andresen, M. A. (2014). The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. Cartography and Geographic Information Science, 42(2), 112-121. <https://doi.org/10.1080/15230406.2014.905756>

Nath, S. V. (2006). Crime pattern detection using data mining. 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops. <https://doi.org/10.1109/wi-iatw.2006.55>

Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from Twitter posts. Social Computing, Behavioral - Cultural Modeling and Prediction, 231-238. https://doi.org/10.1007/978-3-642-29047-3\_28