

Oakland County Data Warehouse White Paper Using Predictive Analysis with Crime Data

Predictive Analysis is a technique that models historical data with assumptive future conditions to predict outcomes or events. With respect to Oakland County's law enforcement community, Predictive Analysis could be a valuable tool for identifying when and where criminal activity will take place.

This white paper examines the processes and tools needed to perform Crime Prediction and Analysis at Oakland County.

Predictive Analysis – What is it?

Predictive Analysis (PA) is the activity of determining what is going to happen in the future by closely scrutinizing historical data and identifying trends, making correlations, and discovering predilections; or more simply put – connecting the dots.

By its own nature, Predictive Analysis is a *proactive* approach rather than traditional business intelligence which is a *reactive* approach. Predictive Analysis uncovers problems before they occur so that you can proactively prevent instead of reactively fix.

Predictive Analysis is successfully used in many industries and business verticals. Most notably, the following business areas have benefited from Predictive Analysis:

- Customer Relationship Management (CRM): to anticipate the buying tendencies of customers and market segments; and
- Fraud Detection: to identify anomalies in credit card activity or insurance claims activity so that fraud and abuse is detected early and minimized.

There are two main methods of Predictive Analysis:

- **Forecasting** – identifying trends with historical data and predicting similar future behavior and events by assuming the trends will continue; and
- **Propensity Analysis** – using mathematical/statistical techniques such as Regression Analysis, Decision Trees, Clustering, Bayesian Networks, and Neural Networks to calculate predilection or predisposition.

Applying Predictive Analysis to Crime

For law enforcement agents, the only thing better than catching the bad guys who commit crimes is catching the bad guys before they actually commit the crime. In order to do that, you have to know **when** and **where** a particular crime is going to occur.

Predictive Analysis can give you the when and where, but there are two fundamental elements that must exist before any Predictive Analysis can take place:

- 1) Relevant historical criminal data must be electronically stored in a data structure that enables analytical processing.
- 2) All data related to crime and criminal activity must be integrated in a common data structure. At a minimum, you need various attributes from Law Enforcement, Jail, Courts, Juvenile Records, and GIS data in the same place.

The most common data structure for intense analytical processing that covers these two fundamental needs is called a **dimensional data warehouse**.

A dimensional data warehouse typically consists of an integration layer (an Operational Data Store (ODS) and/or a data warehouse (DW) database) along with dimensional data marts.

A dimensional data warehouse combines the data from all of the various sources and then organizes the data into facts (statistics, amounts, etc.) and dimensions (time, place, grouping, etc.). With criminal data, the most common fact is simply a count of the occurrence of a crime. However, there are many important dimensions, such as date/time of crime, location of crime, offender name, PACC Code/MCL Code, and many others. Dimensions tell you the **when** and the **where** as well as who, what, and much more.

Thus, if you are able to store years of historical criminal data into a dimensional data model, it will be quite simple to identify trends. For example, you might discover that 90% of car thefts occur between midnight and 2am on weeknights in hotel parking lots. Armed with that information, you can instruct your on-duty law enforcement officers to frequent hotel parking lots more often between midnight and 2am on weeknights.

The same dimensional data structure could also be used with the Propensity Analysis techniques to learn which criminal profiles will be more likely to commit certain crimes or which prevention programs are most effective with would-be criminals.

Some other common Criminal Predictive Analysis functions that would be enabled with a dimensional data warehouse include:

- Crime Pattern Detection – discerning patterns and identifying possible suspects;
- Crime/Suspect Correlation – matching suspects to crimes by examining career criminal files, suspect vehicle files, probation/parole reports, and field interrogation files; and
- Target Profiling – analyzing victim and premises data via GIS.

Criminal Predictive Analysis Tools and Trends

The software tools used to perform Predictive Analysis are classified as **Data Mining** tools. Data Mining tools have been around for many years and some are very well established. The leaders in the Data Mining field are NCR/Teradata, The SAS Institute, and SPSS. Additionally, there are smaller vendors with pre-built Criminal Predictive Analysis applications, such as Sungard's OSSI Crime Analysis Module and ForensicLogic's CrimePoint product. Microsoft's recent announcement about the release of their PerformancePoint Server 2007 (a platform dedicated to Business Intelligence and Analysis) squarely places them as an emerging player in this toolset as well.

One of the emerging trends in Criminal Predictive Analysis is the formation of "Fusion Centers". A Fusion Center is a mechanism for sharing of information and intelligence across local, state, and federal levels of government to improve the ability to fight crime and terrorism. Essentially, the idea is to use the combined financial and human resources from each level of government to develop one, integrated source of information that can be used by all levels. Of course, intelligence products such as Predictive Analysis applications and or data mining tools will need to be used with the integrated data of a Fusion Center to take advantage of its power.

The Department of Homeland Security and Department of Justice have developed guidelines for creating Fusion Centers and they strongly encourage all local and state governments to implement Fusion Centers as quickly as possible.

Criminal Predictive Analysis at Oakland County

Oakland County is in a very good position to implement Criminal Predictive Analysis tools and applications because most of the necessary data has either already been positioned in a dimensional data warehouse structure or is stored electronically and can be moved to a dimensional data warehouse with minimal effort. The main sources of data would be the CLEMIS PrinTrak system, the Criminal Justice Data Warehouse (CJDW), and the Land/GIS databases.

There would likely be additional data needs from external agencies, such as vehicle registration information, driver's license data, OMNI correctional data, and LEIN data from the State of Michigan.

The County would also need to develop or purchase Predictive Analysis software tools and/or pre-built applications to take advantage of its rich data sources.

Conclusion/Recommendation

Criminal Predictive Analysis is most certainly a key element of the next generation of Law Enforcement Reporting and Analysis. The Department of Justice and the Department of Homeland Security has embraced the concept of integrating data in Fusion Centers for the purposes of Criminal Predictive Analysis.

Oakland County is well positioned to be a key member of a regional Fusion Center because of the maturity of its data warehouse program and law enforcement applications. Oakland County should pursue the procurement and usage of Predictive Analysis tools or applications to take advantage of its current data assets and to take the lead on the beginnings of the creation of a regional Fusion Center.