



**Homeland  
Security**

Science and Technology

# TechNote

U.S. Department of Homeland Security



System Assessment and Validation for Emergency Responders

The U.S. Department of Homeland Security (DHS) established the System Assessment and Validation for Emergency Responders (SAVER) Program to assist emergency responders making procurement decisions. Located within the Science and Technology Directorate (S&T) of DHS, the SAVER Program conducts objective assessments and validations on commercially available equipment and systems, and develops knowledge products that provide relevant equipment information to the emergency responder community.

SAVER Program knowledge products provide information on equipment that falls under the categories listed in the DHS Authorized Equipment List (AEL), focusing primarily on two main questions for the emergency responder community: "What equipment is available?" and "How does it perform?" These knowledge products are shared nationally with the responder community, providing a life- and cost-saving asset to DHS, as well as to Federal, state, and local responders.

The SAVER Program is supported by a network of Technical Agents who perform assessment and validation activities.

This TechNote was prepared for the SAVER Program by the National Urban Security Technology Laboratory.

**NUSTL**

For more information on this and other technologies, contact the SAVER Program by e-mail or visit the SAVER website.

E-mail: [saver@hq.dhs.gov](mailto:saver@hq.dhs.gov)

Website: [www.firstresponder.gov/SAVER](http://www.firstresponder.gov/SAVER)

Reference herein to any specific commercial products, processes, or services by trade name, trademark, manufacturer, or otherwise does not constitute or imply its endorsement, recommendation, or favoring by the U.S. Government. Neither the U.S. Government nor any of its employees make any warranty, expressed or implied, including but not limited to the warranties of merchantability and fitness for a particular purpose for any specific commercial product, process, or service referenced herein.

## Data Mining and Analysis Tools

*Data mining and analysis is a process in which usable information is extracted from a database of non-uniform data sources and presented to the user in a clear, legible format. This process streamlines information gathering by allowing the user to gain relevant knowledge from a massive database of information. These tools aid first responders in making strategic decisions quickly and efficiently during emergency events.*

### Technology Description

Data mining and analysis is a type of machine learning algorithm that allows the user to extract patterns and knowledge from a sizable database. Machine learning is a broad classification for algorithms that learn from data and are used to optimize results. Using machine learning, computer scientists and statisticians can perform a wide variety of actions, such as optimizing a model or identifying dependent variables in a data set. Machine learning algorithms are useful when static algorithms that follow a strict set of rules do not apply to the problem.

There are many types of machine learning algorithms, including neural networks and k-nearest neighbors. Neural networks are prominent in the field of data mining. These networks attempt to model the inner workings of the human brain in order to solve complex problems. Using a set of rules called "neurons," a neural network accepts inputs and performs a series of calculations, resulting in an output. These calculations improve and optimize over time, allowing the neural network to adapt to the problem, much like the human brain. Another model used in data mining and analysis is k-nearest neighbors. The k-nearest neighbors algorithm plots points from a dataset and uses the k-nearest points (k being a user-inputted integer) to classify and assess data points. Incomplete data points are extrapolated from the nearest data points, allowing the user to make inferences about missing values.

### Sub-processes in Data Mining and Analysis

The data mining and analysis process is broken down into three sub-processes: data query, data analysis, and data mining.

In the data query process, the user passes specific arguments into the data mining program. The query acts as the interface between the data mining program and the user. The user can specify which analysis and machine learning tools the program should run, along with the dataset that should be used for the process. While this process can be completed by someone with little knowledge of statistics, it is better if the user has some experience working with computer science and machine learning algorithms. Having a designated query builder or data analyst can greatly increase the amount of knowledge and insight gained by this program.

During data analysis, the data mining program takes the user input from the query and analyzes the database. The analysis is done using multiple statistical techniques, such as regression, clustering, and correlation. By doing this, the program can group together different data points and analyze the relationship between them. The k-nearest neighbor algorithm works especially well for this, as it excels at classifying and identifying data.

In data mining, the final sub-process, the program creates a machine learning algorithm that can be run on the dataset. The newly created algorithm will run continuously until the results are optimized. In this case, the optimal result is to display the most relevant information to the user. This algorithm is commonly a neural network or a decision tree. A decision tree accepts several input variables and navigates a series of nodes (points at which decisions are made), eventually returning a prediction for the variable as output. Decision trees can be used in data mining by creating rules that apply to the input (in this case, the database of information) and generate a prediction for the output (the useful information). It runs through this until the result is optimized.

## Availability of Products

Data mining software is unique in that there are many free options on the market available in a wide variety of programming languages. This free, or “open-source” software is available to the general public to use or edit as they see fit. The problem with open-source software is that it requires programming knowledge and a lot of time to optimize for a first responder environment. With this considered, it is not necessarily cheaper to go with free software as opposed to commercial products.

Commercial data mining software is available from several companies in the statistics software market. The purchase of this software usually includes installation by a trained technician. This allows first responders to get the system up and running quickly, and be assured that the product will be supported by an entire company, rather than just the developer of open-source software. The commercial products are very intuitive, and can be used by most people with minimal statistical and computer science experience. However, the software can be expensive: mid-range software can range anywhere from \$2,000 to \$70,000.

## Applications for First Responders

Data mining and analysis tools can have a large positive impact on how first responders handle dangerous situations. In the event of an emergency, responders can use these tools to get a better understanding of exactly what is happening and enhance situational awareness. Data mining tools can extract information from reports and news articles across the Internet, creating a vast database from which to draw relevant information. This enhanced knowledge allows first responders to make superior decisions with regard to equipment usage and response tactics.

In order to fully understand how data mining and analysis tools can help first responders, it is helpful to look at some examples. In the event of a blizzard whiting out an entire town, responders may have difficulty planning their response to a car accident. With data mining and analysis tools, responders can gain information from numerous accident and news reports in order to find the safest route possible and make decisions about what equipment to bring and optimal response strategies.

During Hurricane Sandy, first responders used similar tools to help identify areas where supplies were in dire need. First responders and volunteers analyzed the words and hashtags used in Twitter posts and Instagram photos in order to target their efforts and help expedite recovery.

The effectiveness of data mining and analysis tools is directly related to the amount of time and capital invested. Setting up a large, clean database that contains the most up-to-date information will take time and resources. The data mining system itself, whether using open source or commercial software, will require a significant investment. However, the adoption of this software can provide an invaluable tool to first responders. Numerous software companies have created data mining solutions that, when implemented, can cut down response time, provide knowledge, and save lives.

## References and Resources

- Heaton, Brian. *How Emergency Managers Can Benefit from Big Data*. Government Technology. 23 July 2013. Accessed 8 April 2015.
- SAVER Report. *Data Mining and Analysis Tools Assessment Report*, February 2007.