

June 17, 2018

Course: CIS570 – Business Intelligence
Name: Robert Palumbo
Assignment: Reading Discussions – Week4-Session2
Due Date: Sunday, June 17 @ 11:59pm

Question 1. What do you think is the most prominent application area for data mining? Why?

While there are quite a few tools available for data mining that are in use today and more being created, it is somewhat difficult to identify what is perceived to be the most prominent tool in use. However, I would submit that the R programming language along with the RStudio IDE would be at the top of the list.

R was created by statisticians back in 1992 initially to assist with statistical based data computations etc. While it is a formal programming language, and not the easiest to learn, it has been enhanced over the years to not only provide quick statistical analysis of data, but now includes advanced features such as graphical output, modeling (linear/non-linear), time-series, and clustering.

Further, since it is an *open-source* programming language, it can be easily extended through custom packages submitted by the open-source community typically designed to perform a specific function that is perhaps missing from the base product or adds additional functionality to an existing package etc. At the same time, one can write custom code on the fly to solve a specific problem which with many commercial based products may not be doable at all or requires product support to achieve.

R is particularly useful for performing quick descriptive analysis of data, as well as, data cleansing and formatting. To make things even better, the RStudio IDE (Integrated Development Environment) was created to work together with the R language to create a comprehensive development platform. RStudio is a GUI based interface used to develop and test R scripts and to visualize output either by text or by graphical plots.

Using the tools together allows one to input a dataset, cleanse and prepare the data, create relevant models and run the analysis, and finally assess the results. All in a few *relatively* easy sequence of programmatic steps.

There are many top companies that use R for statistical analysis and modeling including Google, Facebook, Twitter, and Ford to name a few. R is a powerful package in and of itself, is *FREE*, and continues to gain in popularity every day. More and more people and companies are adopting R for its adaptability at being able to solve even the most complex of problems.

Example methods for data-mining using R:

`lm()` is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance

`glm` is used to fit generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution.

`ctree` is used in recursive partitioning for continuous, censored, ordered, nominal and multivariate response variables in a conditional inference framework – building decision trees.

`ts` is used to create time-series objects for time-series analysis

<https://www.r-project.org/about.html>

<https://www.rstudio.com/>

<https://www.datasciencecentral.com/profiles/blogs/list-of-companies-using-r>

<https://www.infoworld.com/article/2940864/application-development/r-programming-language-statistical-data-analysis.html>

<http://www.zentut.com/data-mining/data-mining-processes/>

Question 2: Research and discuss how a specific law enforcement agency uses predictive analytics.

The use of predictive analytics in law enforcement is becoming widespread and helping to assist law enforcers in performing their duties more efficiently and safer, which leads to a better outlook in the fight against crime and criminal activities. One specific use is known as *Predictive Policing*. The goal with this technique is to take a more proactive approach in preventing crimes *before* they occur. The process relies on using historical data (required by predictive analytics) to *predict* potential higher crime areas.

For example, agencies use big-data analytics to determine which areas of a city exhibit higher rates of crime, as well as, the time periods in which a surge in crime may likely occur based off those same analytics. From this, agencies can direct more patrols to canvas the affected areas thereby providing a more active presence with the hope to mitigate potential criminal acts.

The Kansas City Police Department uses predictive analytics to identify areas in the city that have shown an increase in criminal activity, and to further, identify parolees that are known to be living or within the affected areas and contacting those individuals to verify their status and to let them know they are under watch. The hope is that the department can intervene and assist these parolees to keep them on the right path as it is well documented that convicted offenders often fall back into the same patterns once release from custody.

The Baltimore County Police Department used analytics to apprehend an individual that was responsible for a series of convenience store robberies. The police only had the locations of each robbery and description of a suspected vehicle used by the perpetrator. They then employed geo-mapping techniques to identify an origin point and from that, analyzed the streets surrounding the crime area map to predict which street was likely used for all associated crimes. This led to the continual surveillance of the identified street which led to the eventual apprehension of the suspect. Using analytics, the department could focus in on a specific location versus having to monitor and patrol all of the other surrounding convenience stores that may or may not be the next target.

The FBI cybercrime unit uses this type analytics to predict criminal acts of cyber security to identify where a breach might occur or to circumvent potential leaks or hacks before they occur.

A side note to this topic is that critics would argue that predictive models are based on past data and that crimes may be missed that do not fall into the predicted patterns – due to the fact that not all crimes are reported. Still, I believe the models become more and more accurate each day as additional data is added and the model are retrained to use that data.

<https://insidebigdata.com/2017/09/08/big-data-law-enforcement-rise-predictive-policing/>

<http://kansascitynova.org/>

<https://www.wsj.com/articles/is-predictive-policing-the-law-enforcement-tactic-of-the-future-1461550190>