**BLOG 1**

**Data Analytics in Infrastructure monitoring**

My current team works on infrastructure monitoring. We manage a tool that monitors Windows servers globally. The initial use of the tool was reactive, as teams were alerted for issues after a system was already in a declining state. That past few years we have begun digging into the event, performance and data collections in an attempt to correct issues with automated task and scripts to relieve performance issues. Our next phase is the use data analytics to predict issues before it degrades to a point that action has to be taken to correct.

Statistical predictive analysis methods would be useful for this goal. Specifically, multiple regression analysis. The objective would be to predict which combination of threshold metrics can predict a system performance degradation before it gets to an unstable state.

**Approach**

The approach is to collect as many variables as possible for specific categories of servers. We would collect data for server that are used for specific functions like Active Directory, Exchange, SQL server. The predictor variable would be the performance metrics in time and response variable would be weather a trouble ticket is generated.

Our application already collects the performance data and discovers machine functions by install paths, registry keys and other methods. We are beginning to collect ticket information for the analysis. With these 2 datasets we can use a time based GLM to predict outages and system troubles.

**BLOG 2**

**Data Science in Human Nature**

“Wisdom doesn’t necessarily come age, sometimes age just shows up all by itself”

We try to understand our place in time inside the mounds of history books that have been etched into the fabric of global networks worldwide. With this limitless connected data, can we now correlate the advancement of technology with moral and social advancement or is technology moving faster than we can reflect and learn? Can data science help us learn our past and direct a better course or are human nature behaviors stagnant, leaving us doomed to overlooked lessons of history and repeat the errors of our nature.

Learning and understanding human nature and behaviors is the holy grail of social sciences. Understanding why we do what we do, at a specific point of a lifetime can vary from person to person, but in the aggregate can we map enough qualitative and quantitative data to make sense of our unpredictable predictability?

**Modelling**

Understanding human nature can be an elusive task for data science. We can make hundreds of binary decisions a day, with each decision setting us on a new course the following moment. Basic linear modeling would probably not suffice, there are many unpredictable variables and confounding factors. Forward based and backward based models would not suffice either. Automated step wise models would probably perform fractionally better, but can fall short. If we add nonparametric regression to GLM with time series data, it may allow enough flexibility for predictions.

The biggest factor for a successfully modeling human nature will be to narrow down the scope as much as possible. Focusing the predictors on things that affect our basic needs of food, clothes and shelter may be a good approach to determine the algorithm of our nature.

**The Challenges**

This NSI article (<https://nsiteam.com/social-science-modeling-visualization/> ) discusses some of the core challenges of modeling for social sciences. Some challenges include:

* The need for better scientific theories
* The need to clarify and scope the core question
* The lack of strong datasets for social science research
* The restriction of some of the social science datasets and private classification of data

The NSI article focuses on social sciences for national security, but I believe it may still apply broadly, although it may be dated. The article was posted in 2008 and technology has advanced the past decade. Since then, more information has been made available with the expansion of social networks, the Internet of Things and Wearable technology. These devices and networks have made us more connected, and for a data social scientist, this plethora of data may have the keys to help us learn and reflect about our place in time now and beyond.

**BLOG 3**

**Investing**

Modeling in investing has become the norm in financial institutions. The employees that do the analytics are called quants or quantitative analysts. These quantitative financial wizards use mathematical modeling to analyze markets trends. The analysts go through massive datasets to determine the direction of a security or market. The amount of data used to get sense of direction is mind boggling.

**Modeling**

Below are some of the model quants use, based on information in Wikipedia:

(<https://en.wikipedia.org/wiki/Quantitative_analysis_(finance)>)

* [Finite difference method](https://en.wikipedia.org/wiki/Finite_difference_method) – used to solve [partial differential equations](https://en.wikipedia.org/wiki/Partial_differential_equation);
* [Monte Carlo method](https://en.wikipedia.org/wiki/Monte_Carlo_method) – Also used to solve [partial differential equations](https://en.wikipedia.org/wiki/Partial_differential_equation), but [Monte Carlo simulation](https://en.wikipedia.org/wiki/Monte_Carlo_simulation) is also common in risk management;
* [Ordinary least squares](https://en.wikipedia.org/wiki/Ordinary_least_squares) – used to estimate parameters in [statistical regression analysis](https://en.wikipedia.org/wiki/Regression_analysis);
* [Spline interpolation](https://en.wikipedia.org/wiki/Spline_interpolation) – used to interpolate values from [spot and forward interest rates curves](https://en.wikipedia.org/wiki/Yield_curve), and [volatility smiles](https://en.wikipedia.org/wiki/Volatility_smile);
* [Bisection](https://en.wikipedia.org/wiki/Bisection_method), [Newton](https://en.wikipedia.org/wiki/Newton_method), and [Secant methods](https://en.wikipedia.org/wiki/Secant_method) – used to find the [roots](https://en.wikipedia.org/wiki/Root_of_a_function), [maxima and minima](https://en.wikipedia.org/wiki/Maxima_and_minima) of functions (e.g. [internal rate of return](https://en.wikipedia.org/wiki/Internal_rate_of_return).)

There are 2 models on the list that we discussed this semester, ordinary least squares and spline interpolation. In the ordinary least square quants try to find the quadratic formula that estimate points nearest to the regression line that minimizes mean and variance values. In spline interpolation quants try to determine the polynomial formula that show where points, or in this case knots, have a change in direction.

**The Challenges of Quantitative vs Qualitative approach**

The challenges are many. If this type of prediction was easy, we would all be millionaires. Many times quantitative data is not enough to predict markets, there are also qualitative data that need to be considered. In this global economy bad weather, an unstable political environment, wars and other events in any part of the world can impact a market’s direction. For every analyst that predicts a down market there are 10 others that predict an up market.

With that said, the idea behind the quantitative approach is that the market will ebb, flow and react to the numbers, news and event on their own. Therefore, the plot behavior is the market reaction to news, earnings and anticipation of future events. In theory, if you keep your eyes on shift of direction of the plots, you should see the shift in direction of the markets.

**BLOG 4**

**Sports Analytics**

There has been a shift in sports analytics the past few decades. The impact has been positive for teams who are willing to invest into analytics that can provide the models to predict players impact on their sports. The objective is to minimize cost by trying to find players that are “diamonds in the ruff”, those impact players that are a value-add with minimum cost.

In the early days of baseball, the original stats of batting average and runs was used to determine how a player can help a team win games. In the 1990’s Billy Beane began to use a sabermetric method that was first researched by Earnshaw Cook decades before. This method analyzed all statistical data from the game, not just hitting, but pitching and fielding.

**Methods**

The below methods are from Wikipedia link (<https://en.wikipedia.org/wiki/Sabermetrics>)

|  |
| --- |
| Sabermetrics methods are generally used for three purposes:  1. To compare key performances among certain specific players under realistic data conditions. The evaluation of past performance of a player enables an analytic overview. The comparison of this data between players can help one understand key points such as their market values. In that way, the role and the salary that should be given to that player can be defined.  2. To provide prediction of future performance of a given player or a team. When past data is available about the performance of a team or a specific player, Sabermetrics can be used to predict the average future performances for the next season. Thus, a prediction can be made with a certain probability about the number of wins and losses.  3. To provide a useful function of the player's contributions to his team. When analyzing data, one is able to understand the contributions a player makes to the success/failure of his team. Given that correlation, we can sign or release players with certain characteristics. |

For each of these methods teams can use regression analysis and ordinary least squares with likelihood ratio. With MLB new era of analytics, teams are discovering more ways to use statistics to their advantage. For example, Value over replacement Player (VORP) measures how a player contributes to his team in comparison to a fake replacement player that performs below average. Win above replacement (WAR) also compares to a fake replacement player, but determines how many wins the player contribute to the team.

**Changes in the Game**

The new analytics has revamped hiring of managers and coaches in Major League Baseball. Hiring is now based on a coach’s ability to use analytics during real game situations. It is also changing the way teams position players on the fields for defensive plays. Players are learning that their swing launch angle is important and can be the difference between a double and a home run. Analytics teams can also help a player focus on their training and regiment. Even when things don’t go the player’s way, analytics can explain that even if results don’t show at this moment, if they keep at it, the trend show they are about turn a corner.

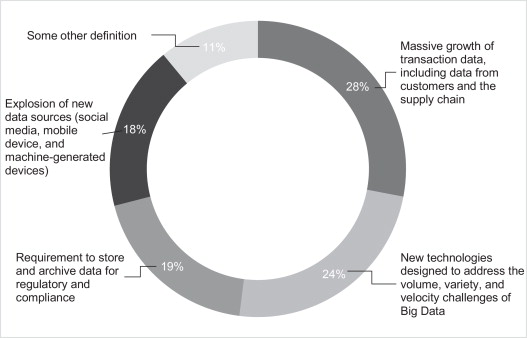
The win for teams can be a win for fans if teams allow their fans time to adjust and accept changes. The teams must find a pace that works. If they move too slow, they risk losing a title, if they move too quickly, their changes can result in losing fans which in turn becomes a loss in revenue.

**BLOG 5**

**Retail Analytics**

In 2012, Target mailed coupons to a high school girl for baby clothes. As a result of her previous purchases, Target was able to predict that the girl was pregnant before her parents became aware. The details on how Target was able to predict this was all over the news. This was the public’s first glimpse of what was to become the norm.

Today if you do a search in Netflix or Amazon, you will get recommendations for similar movies and products. Siri, Chrome, Alexa devices record conversations and can use speech to text technology with linguistics to analyze discussions and make recommendations. Social media collects data on likes and discussions, which can be used by advertisers to try to predict your future purchases.



*Reference:* [*https://www.sciencedirect.com/science/article/pii/S0268401214001066*](https://www.sciencedirect.com/science/article/pii/S0268401214001066)

**Methods**

The days of mailing surveys to collect consumer data are long gone. Today that same information is everywhere and it requires text, video and audio tools to analyze. Predictive analysis requires some varying methods to analyze this era of Big Data. Some methods include correlation matrixes, sentiment analysis, natural language processing (NLP) and others to name a few.

**Challenges**

Retail analytics try to make predictions based on the aggregate. However, there are some privacy concerns on how the data is used and weather an organization can protect an individual’s identity. The concern is, if the identity security is breached, the analytics can be inaccurately interpreted and/or be used for discrimination or manipulation. There is also the issue of data being housed for a lifetime. How long should the data be housed before being destroyed and can we count on businesses to follow through with data cleanup? There are also ethical concerns. Just because a company can do something that can boost revenue and cut cost based on the analytics, should they do it although it may adversely affect lives.

Predictive analysis has become a boon for businesses and can be a win for consumers. The theory is prices will go down as businesses cut cost with more efficient targeting. This in turn lowers cost on consumer good. Even with this win-win, laws must be in place to avoid any unethical uses of this personal information.

IDEAS

* **Infrastructure Analytics**
* Who’s Idea is it anyway
* **Retail**
* **Sports analytics**
* Population Trends
* Predicting Next Superpower
* Global Warfare predictions
* **Our place in time Have technological Advances impacted Human Nature**
* **Investing**