Saving Lives at Work:

The Who, What, Where, Why and How of Using Predictive Analytics in Workplace Safety

A Predictive Solutions White Paper

Predictive SOLUTIONS
An Industrial Scientific Company

Research conducted by teams at Predictive Solutions Corporation and Carnegie Mellon University has shown that workplace injuries can be predicted, by computers, with accuracy rates as high as 97%.

Do you believe that workplace injuries can be predicted? Regardless, would you believe that machines are better at predicting workplace injuries than humans? Research conducted by teams at Predictive Solutions Corporation and Carnegie Mellon University has shown that workplace injuries can be predicted, by computers, with accuracy rates as high as 97%.[1]

Beyond the field of safety, there is a mountain of research results to suggest that, hands down (pun intended), machines are better at predicting future outcomes than humans. This has been proven with regard to parole decisions[2], medical diagnoses[3], and business supply chain optimization[4], just to name a few.

As a result, according to many researchers, including Andrew McAfee in the Harvard Business Review, "The practical conclusion is that we should turn many of our decisions, predictions, diagnoses, and judgments—both the trivial and the consequential—over to the algorithms. There's just no controversy any more about whether doing so will give us better results."[5]

To the contrary, there are many who believe that there is still a huge role for humans to play in predicting, but certainly preventing, workplace injuries. Thomas Davenport, one of the foremost experts on data analytics, said "there are plenty of places where [human] intuition is still relevant."[6]

Regardless, along with various human-based measures, employing machine learning algorithms to predict workplace safety incidents is absolutely critical in a highly evolved 21st century safety program. There is no debate that the use of predictive algorithms has been successful in other business applications like credit scoring (FICO), Internet search results (Google), and shopping (Netflix and Amazon), and now they are proven in safety.

This paper will take a comprehensive look at this topic by answering the following questions:

- Who is employing predictive analytics in safety?
- Why are companies employing predictive analytics in safety?
- Why are companies struggling to deploy predictive analytics programs on their own?
- What is predictive analytics as applied to safety?
- How is predictive analytics applied in safety programs?
- Where is safety prediction headed in the future?

[1] http://www.predictivesolutions.com/lp/four-safety-truths-reduce-workplace-injuries/

[2] https://books.google.com/books?id=J6iq_khf5HkC&pg=PA89&lpg=PA89&dq=dawes+parole+boards&source=bl&ots=1bmZ3Q3RF8&sig=Xt8texpgV2wSY369d3FJ3xjm-Ho&hl=en&sa=X&ei=WhWeUoKqAuWvsQTYzYCoAg#v=onepage&q=dawes%20parole%20boards&f=false

[3] http://andrewmcafee.org/2011/12/using-computers-to-cure-medicines-pathologies/

[4] http://www.chrissnijders.com/me/files/3cafc909ea68ba79c728d7d06d7a34ff-5.html

[5] https://hbr.org/2013/12/big-datas-biggest-challenge-convincing-people-not-to-trust-their-judgment/

[6] https://hbr.org/2013/12/big-data-and-the-role-of-intuition/

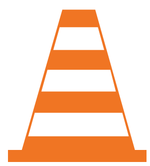
Predictive SOLUTIONS
AN INDUSTRIAL SCIENTIFIC COMPANY

Who is Employing Predictive Analytics in Safety?

Nearly every industry and companies within have deployed some level of advanced or predictive analytics in safety. What started out as a niche practice is growing quickly. Some of the first applications of this methodology occurred in the construction industry around 2004, but it has now been applied in most industries including general and specialized manufacturing, oil and gas, energy and utilities, retail, aerospace, automotive, food and beverage, and even the public sector and universities. The application is also global with practitioners on nearly every continent. The results that companies are achieving are impressive:

Top 5 public U.S. university saved over $20 million in insurance fees across four years of construction projects

A Fortune 150 energy company reduced its injury rate by 67% within 18 months



A top 20 construction company achieved significant safety improvements including 90% of worksites experiencing no lost-time incidents

A Fortune 150 manufacturer reduced its lost work day rate by 97% within one year

The majority of the companies that are employing predictive analytics already have a strong safety culture and impressive safety results, which is a great segue to the next question.

Why are Companies Employing Predictive Analytics in Safety?

The use of predictive analytics in safety isn't a good fit for every company. Though some exceptions exist, companies with high levels of injuries generally do not adopt the use of predictive analytics. Their safety programs simply haven't evolved yet to a place where they can employ the strategy effectively. Instead, these companies usually get greater benefit from analyzing their lagging indicators, doing root cause analyses of their incidents, and then using that information to lower their injury rate. Though this is a more reactive approach, it can be effective. Unfortunately for their employees, there is a great deal of "low hanging fruit" in the form of injuries that can be analyzed to reduce risk.

Companies that employ predictive analytics in safety have usually evolved past this stage in their safety culture. They may have employed a safety culture change program, or invested in the development of a safety management system, or started collecting leading indicator data in the form of near misses or safety observations. They might even have done all of these things. Regardless, they are generally companies that have made a commitment to take a proactive, rather than a reactive approach to safety.

Ultimately, there are two main reasons why companies employ predictive analytics in safety:

1. They have a continuous improvement safety culture and are always on the lookout for new 21st century methodologies that work, and/or

2. To break through a safety performance plateau when more traditional safety strategies have become less effective.

Why Do Companies Struggle to Deploy Predictive Analytics Programs on Their Own?

Regardless of why a company chooses to adopt a predictive analytics strategy, deploying a predictive analytics program can be challenging.

First, some companies struggle because they are not collecting the right type of data. As detailed later in this paper, in order to predict injuries, both target variable data (e.g. actual injuries) and input variables are required. Many companies have not made the investments in time (and sometimes technology) to gather the type of input data required to drive predictive models. This data might include near-misses, safety observations, and/or training records. Often, safety professionals are ready to make this investment, but need help in identifying the right type of data to collect in order to produce accurate models. These companies simply don't know where to begin.

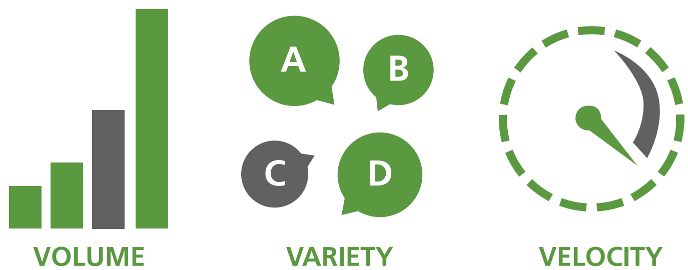
However, most companies have the exact opposite problem. They are collecting so much data that they're drowning in it and can't make sense of it.

According to IBM, 2.5 quintillion bytes of data are created daily and 90% of the data in the world was created in the last two years.

And it's no wonder since we are in the era of Big Data[7]. According to IBM, 2.5 quintillion[8] bytes of data are created daily and 90% of the data in the world was created in the last two years.

The collection of safety data has similarly skyrocketed. In July 2015 alone, one company recorded more than 200,000 safety observations. At this rate, they will collect over 2.4 million in a single year! This is driven by more data collectors, but also by improved technology. The use of smartphones and tablets has driven a massive increase in the amount of safety data collected at work.

With regard to Big Data challenges in safety, there are "The Three V's":



- Volume – As noted above, overall volume is increasing exponentially.
- Variety – As safety strategies broaden and data collection technologies improve, the types of data collected are expanding immensely (e.g. using a Fitbit to collect physiological data).
- Velocity – As data storage, transfer, and processing technologies improve, the speed at which data can be actionable increases tremendously.

But safety professionals often find that they can't get a return on investment from this huge data collection effort. Initially, because the data resides in so many different systems, they struggle to extract it and get it all in one place. This usually requires IT people and resources that are either lean across most businesses or are rarely at the beck and call of the safety function.

Even after overcoming this first challenge, it then can be difficult to normalize or link the disparate data sets in any meaningful way in order to prepare it for analysis. This is often called "data wrangling"[9] and requires significant data architecture expertise. Finally, if these first two hurdles can be overcome, the resulting massive data set is just too large and too complex for simple analysis tools, let alone the human brain. Most safety professionals, or even their respective companies, do not have access to the complex systems and expertise required to analyze such Big Data sets.

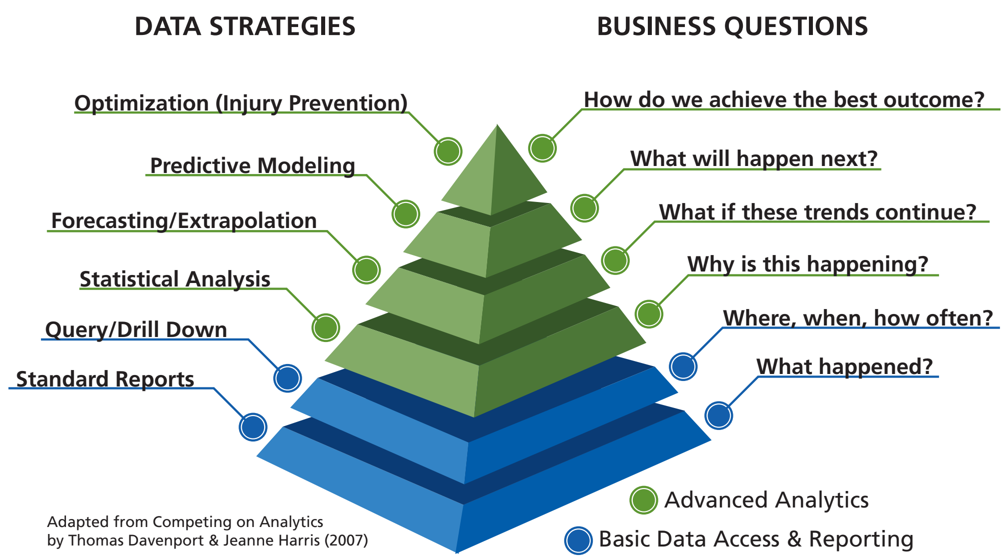
[7] https://en.wikipedia.org/wiki/Big_data
[8] That's a 25 followed by 18 zeros or 25,000,000,000,000,000,000
[9] https://en.wikipedia.org/wiki/Data_wrangling

But do not despair. Many of the technological advancements that contribute to these Big Data challenges also contribute to their solutions.

First, the exponentially increasing computational power resulting from Moore's Law[10] allows for data analysis systems that are more robust than ever. This allows for advanced and predictive analytics capabilities that allow us to answer, with high confidence levels, ever more complex business questions.

This is represented in the graphic below that was adapted from Thomas Davenport's seminal book "Competing on Analytics."

The Safety Analytics Pyramid



The above paradigm, referred to as the "Safety Analytics Pyramid," suggests that as data is collected, various "data strategies" (on the left-hand side of the graphic) can be executed in order to answer various "business questions" (on the right-hand side of the graphic). Moving up the pyramid, more advanced data strategies are required to answer more difficult, yet also more meaningful, business questions.

This paradigm suggests that companies who only have access to basic data capabilities (e.g. basic spreadsheet functionality) will be limited to only being able to answer the basic business questions at the bottom of the pyramid – questions like "what happened" and "where, when, and how often?" But the answers to these questions only yield lagging indicators and don't say anything about the future. For companies that want to be proactive versus reactive in their safety program, this is not enough.

Leaders don't want reports, they want answers! In order to move up the analytics pyramid and answer more strategic and proactive business questions – like "why is this happening" and "what if these trends continue" – a company needs access to even more robust data strategy expertise and systems.

[10] https://en.wikipedia.org/wiki/Moore%27s_law

If they want to answer the penultimate question of "what will happen next," with a high confidence level, Tom Davenport says a predictive model is needed. In the realm of safety, leaders are trying to answer the question "where and when will our next incident occur?" Once this question can be answered, a company can turn to the effort of "achieving the best outcome" or, in this case, preventing that injury from occurring by optimizing their response with countermeasures to reduce risk.

So, just as is the case in other business functions, technology tools are available to safety professionals to make accurate future predictions about safety.

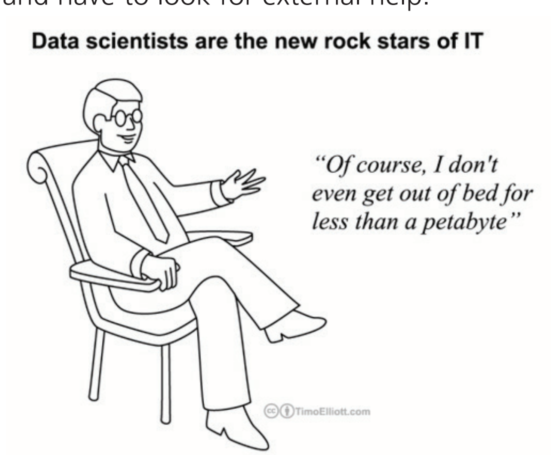
So, just as is the case in other business functions, technology tools are available to safety professionals to make accurate future predictions about safety. Big Data safety problem solved, correct? Not so fast. Those systems don't run themselves – not yet anyway.

A unique set of skills and expertise is required to make these systems hum. First, someone is needed to interpret how predictive models are working so that companies can be directed as to which data sets (input variables) need to be collected to feed the models to get the desired answers. In addition, someone is needed to wrangle that data and get it into a structure such that a computer can do something with it. Next, someone needs to configure, load, operate, tune, and then interpret the output from machine learning predictive modeling software that use their own languages, their own parameters, and have their own nuances.

Such a person who can do all these things must be very technical and scientific in nature. However, to be truly effective, they also need to be able to relate the findings of these complex models to safety professionals and business leaders in such a way that they can first understand them, and then take action on them. In this case – to prevent predicted injuries from occurring.

This is no small task. Regardless, these people exist, and they are called Data Scientists. Forbes magazine recently called the role the "sexiest job of the 21st century." According to IBM, a Data Scientist is "part analyst, part artist…with the goal of discovering a previously hidden insight, which in turn can provide a competitive advantage or address a pressing business problem."[11]

Data Scientists, and the tools they employ, are in short supply generally[12], but are an even rarer breed in the realm of safety. This is why companies often struggle to address their Big Data issues themselves, and have to look for external help.



[11] http://www-01.ibm.com/software/data/infosphere/data-scientist/
[12] http://www.forbes.com/sites/gilpress/2015/04/30/the-supply-and-demand-of-data-scientists-what-the-surveys-say/

Predictive SOLUTIONS
An Industrial Scientific Company

What is Predictive Analytics as Applied to Safety?

In short, predictive analytics applied to safety is the practice of collecting data, and then moving to the top of the Safety Analytics Pyramid to predict where and when future safety incidents are most likely to occur so that they can be prevented. This practice goes well beyond developing lagging indicators from data, and even goes beyond creating leading indicators. From an analytics perspective, it is the penultimate evolution of a 21st century safety program; the ultimate step being optimization.

This practice employs analytics that go far beyond what the human brain, or even basic analytics systems, can handle.

Further, this practice employs analytics that go far beyond what the human brain, or even basic analytics systems, can handle. Because of the complex characteristics of the aforementioned Big Data sets that are used in safety, machine-learning algorithms have proven to be the best tools to create safety predictive models. They can find patterns, correlations, and hidden meaning in what appear to be opaque Big Data sets.

The process of creating these predictive models, managed by an experienced Data Scientist, goes something like this:

- Historic Big Data sets are combined and then "wrangled" into a structured framework. This structure must consist of a dependent or target variable (the variable intended for prediction which, in this case, is workplace safety incidents) that can be linked (usually by time and location) to independent or input variables (the variables being used to predict, such as safety inspections, observations, training logs, etc.).[13]
- This structured data set is then fed to high-powered machine learning software systems which allow a computer to learn how the independent and dependent variables correlate and in the process, develop a model based on patterns it recognizes in the interaction of the variables. In almost all instances, these data sets need to be sufficiently large to allow for thousands of machine learning cycles.
- Once the machine has done enough learning and has developed a model, that model needs to be tested against a test data set of the exact same structure as the learning data set.

At the end of this process, only two outcomes are possible; accurate models or inaccurate models.

With inaccurate models, it is possible to repeat the process until more accurate results are achieved. This usually requires fine-tuning of the modeling strategies, data sets, and other nuanced factors by a Data Scientist. If better results can't be achieved, then the data is simply deemed to not be predictive, which does happen occasionally.

Though accurate models can be deployed into production, they need to be constantly monitored, tested, and refreshed as changes to the working environment are almost constantly occurring.

[13] Many data scientists contend that this part of the process is usually the least understood, least appreciated, and most overlooked. Yet, at the same time, it is the most time consuming, difficult, and requires the most human intervention. Regardless, if not executed properly, the next steps cannot even be attempted.

How is Predictive Analytics Applied in Safety Programs?

Organizations that want to take a proactive, 21st century approach to their safety program add predictive analytics as an additional component to their existing, more traditional safety programs. Sometimes they are imbedded within other safety software systems, or they can be deployed as independent, stand-alone, customized solutions.

To explore this further, following is a real-life example of predictive analytics applied to safety.

On February 23, 2013, an employee in Hazard, Kentucky lacerated his arm with a box cutter while trying to unpack some shipping crates. Ultimately he was fine, but he had to go to the hospital and receive stitches.

By most traditional safety measures, this location was deemed to be "safe." In fact, it had not had a recordable safety incident for nine years to the month leading up to the laceration. This location had a strong safety culture built on strong traditional safety strategies including a behavior-based observation program, a robust safety training schedule, and an engaged leadership team that used bottom and middle-of-the-pyramid analytics techniques to help manage safety and create leading indicators of risk.

Regardless, no human being saw this incident coming, not the employee, the site safety manager, or the plant manager. However, a machine did.

Regardless, no human being saw this incident coming, not the employee, the site safety manager, or the plant manager. However, a machine did.

On February 1, 22 days before this incident, a predictive model flagged this Kentucky location as being at high risk of having increased safety incidents over the next 30 days. While no human predicted this incident, the machine-learning computer model did.

Around this same time period, this location's business unit had received a total of four "red flag" predictions of increased incidents. In three of those instances, injuries occurred. At the one location that did not have an injury, the company was unsure if the model had made a bad prediction, or if the action they had taken to prevent the injury had worked.

Regardless, after seeing the accuracy of the models, they decided to adopt a new strategy in deploying predictive analytics to safety. Rather than, as Andrew McAfee might suggest, abandoning the traditional and mainly human-based safety practices that had helped them achieve a world-class incident rate, they employed the 'Tom Davenport strategy' and simply augmented them with machine-based predictive analytics.

1. First, they review the feedback from the model and incorporate that into their response.[14]
2. Second, they drill back into their data, moving down the Safety Analytics Pyramid, and come to human-based conclusions about their risk. They look at data elements such as safety observations, training records, productivity schedules, etc.
3. Third, they go out to the flagged worksite and talk directly to the managers and front-line employees about the risks that they perceive to be happening at that jobsite.
4. Finally, based on all of this information, they have a safety stand-down to talk about the specific risks they've uncovered, and how to effectively manage them.

[14] Though the models can generally only identify correlation rather than causation, some causal relationships have been found as a result of the modeling process. Regardless, since a detailed discussion of correlation versus causation is beyond the scope of this paper, additional information on the topic can be found here - https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation.

Adding predictive analytics into its existing safety program has helped this business unit drive its incident rate down, but just as importantly, reduce its cost per incident from $24,000 to just $1,205 – a 95% improvement – in three years. Now, rather than just reacting to incidents, they can be proactive by predicting and preventing them.

Where is Safety Prediction Headed in the Future?

A few key trends will likely impact the practice of predictive analytics in safety into the future.

First, as Moore's Law continues to drive machine innovation, more robust analysis tools will be developed leading to better models. While the safety function will most likely not lead this effort, safety professionals will be the beneficiary of the advancements made by the pervasive application of predictive analytics on Wall Street, in other business functions, in medicine, etc.

Second, access to more data will drive better models. In a recent R&D effort, it was found that adding just one new type of data to an existing model resulted in an increase in accuracy of 80%. Adding a second new data type yielded another 20% increase in accuracy. Generally, more data results in better models.

However, there is a cost to collecting this data. Companies have to invest their people's time, or their dollars, or both to collect data to feed the models. As technology improves, this cost decreases, but not to zero. This has to be considered in the endless pursuit of more data.

Third, the need for industry-specific and even company-specific models will increase as safety becomes more specialized. Historically, safety has been a fairly general practice across industries. As a result, the application of predictive models that are fairly general has been successful, partly because the data available has been relatively similar. However, that is changing as industries and companies continue to adopt safety practices that are much more specialized resulting in more unique data sets within industries and companies. Some of the highest performing models to date were built for individual companies based on their specific data sets alone.

However, this point is often over stated. First, for many companies, a specific model based just on their data is simply not possible because most companies don't have enough data to drive the machine learning stage of the process. As stated, large data sets are required to drive thousands of iterations of learning cycles by computers in order to build models. Often, data from multiple companies must be aggregated and thus that data needs to be fairly general and uniform. Second, almost every company likes to think of it and its data as being unique, but that is seldom the case, especially with regard to safety.

Finally, yet probably the most vexing trend for the future, is the need to re-think how the effectiveness of predictive models in safety is measured. What role should model accuracy actually play in measuring its effectiveness? For example, what if a model, that when tested performed at an accuracy rate of 99%, predicted 10 injuries at a company over the next 30 days? Next, the company took action, based on those predictions, to reduce their risk of actually incurring those 10 injuries. Then, over the next 30 days, no injuries occurred. Was the model 100% effective because there were no injuries or 100% inaccurate because there were no injuries? If you believe the overall goal of a safety strategy is to completely eliminate injuries, then the former is the right measure. The bad news, depending on perspective, is that predicting injuries is easier than preventing them at the moment, so we don't have to worry about solving this issue in the short term. There are still far too many injuries left to predict, let alone prevent.

Eventually, we will need to stop predicting unsafe outcomes and start predicting safe outcomes. To do this, we will need to collect data not on what predicts injuries, but what predicts NO injuries. In doing so we will need to move to the very top of the Safety Analytics Pyramid and model the optimization of the deployment of the right mix and quantity of all the safety strategies available to us.

Conclusion

Whether machines or humans are better at predicting and preventing workplace injuries is the wrong question to answer. The right question to answer is "where and when will your next injury occur?" Can you answer that question with either a human or a machine?

Research has proven that in safety, just as in other business functions, machines can predict future outcomes with incredible accuracy. Company results have also proven that if machines can predict future safety incidents, humans can prevent them.

In a recent report, the research firm Verdantix noted that 88% of CEOs believe that environmental, health and safety performance impacts financial performance. As discussed earlier, these CEOs don't want reports, they want answers! We can now provide those answers by employing the same advanced and predictive analytics techniques that have been used successfully in other business functions to the field of safety.

1 Life Way ▪ Pittsburgh, PA 15205-2500 ▪ USA
Tel: 412.809.1888 ▪ 1.800.991.3262 ▪ Fax: 412.788.8353 ▪ info@predictivesolutions.com
www.predictivesolutions.com