A HOW TO GUIDE TO PREDICTIVE ANALYTICS
A HOW TO GUIDE TO PREDICTIVE ANALYTICS
Chapter 1: The Promise of Predictive Analytics

Chapter 2: Data End Points

Chapter 3: Storing and Managing Data

Chapter 4: Policy

Chapter 5: Visualizing Data

Chapter 6: Use Cases

Why ViON?
1: THE PROMISE OF PREDICTIVE ANALYTICS
Today’s advanced predictive analytics reverses the historical paradigm of how we interact with data. In the past, data collected by businesses, governments and organizations has been analyzed in a forensic manner. In other words, the analysis is always looking back at what has occurred, without much capability to use the historical patterns in a predictive way. Make no mistake, this approach had some real benefits, but without the more sophisticated big data and analytic capabilities available today, the predictive element has been more guesswork than science until now.

Today, as we see dramatic advances in the algorithms that model patterns to examine large data sets, we have the unique (and heretofore unimaginable capability) to use data to accurately predict future events.

The promise of this capability delivers benefits in every imaginable category, from resource allocation in law enforcement and defense, and optimized energy use in so-called smart cities, to re-admittance avoidance in healthcare, and even retail settings tuned to the highest fidelity for shoppers likely to walk in specific stores at specific times looking for specific things.

Today, predictive analytics genuinely holds out the prospect for a better world where people do not scurry around in a defensive, reactive mode, trying to solve problems after they occur. We are now predicting the likelihood of events, and allocating the appropriate and proportionate resources in advance of their occurrence.
2: DATA END POINTS
The data end points that can be accessed by predictive analytics solutions are only limited by a user’s imagination. For instance, in healthcare big data applications, predictive analytics can extract – and make predictive sense of – such granular data as caregivers’ appointment records, doctor’s notes from an ER visit, medications dispensed by multiple pharmacies that don’t even talk to each other, 911 records, diet and family history, to name just a few categories. When blended together, these data points can offer insights and decision-making capability that are critical for improving care, decreasing costs and improving clinical outcomes.

If predictive analytics can manage problems like personalized medicine, imagine what the analytical power can bring to issues such as national defense, cyber attacks or threat assessments. The power exists to bring analytical power not only to the everyday lives of people, but to global enterprises that ensure the safety of nations.

The data that needs to be subjected to a predictive analytic examination largely exists. In fact, it sounds like an urban myth, but it is true that 90% of the world’s data has been created in the last few years. With the acceleration of data collection, whether it’s medical records and X-ray’s of your children’s sprained ankle, or IP addresses of bad cyber terrorists on a global stage, we will see the need for data storage and data structuring to expand exponentially at an ever-increasing rate. It is not unimaginable that someday soon 90% of the world’s data will have been created in the last few months.

For a data analytics solution to improve accuracy over time, it must build on the previous results and learn from that. The results of previous analysis, become part of the data pool for future analytics and over time, a cycle develops where the data analytics solution learns from itself.
3: Storing and Managing Data
Storing and managing the amount of data that can be subject to a predictive analytic examination is no small task. But it is not the storage devices themselves that are the challenge. In fact, the vast majority of data that is has been created, and will be created in the near-term, is largely “unstructured.” By way of comparison, structured data organized neatly in databases, such as the data used by banks and financial institutions. Unstructured data is data emulating from emails, video feeds, photography, electronic images, or data collected by some of the millions, if not billions, of sensors that are proliferating in our built environments. It is estimated that 80-90% of data is unstructured.

Candidly, the challenge for predictive analytics is not creating more and more powerful algorithms to interrogate the databases. These powerful predictive analytic solutions exist today, and they will be improved over time at astoundingly rapid rates. The challenge for predictive analytics is to structure the data that is presently unstructured, and to access storage and retrieval methods that can access on-prem and cloud databases, while reaching across to allied or external, even third party databases and access that data as if it were held in a native device. This genuine challenge can best be illustrated by looking at the task of managing so-called “smart city” data. “Smart city” data is the data collected by such things as traffic sensors, video monitors, and the flow patterns of pedestrians, as well as facial recognition, individualized purchase histories, and even criminal records. This data can be gathered in multiple forms from thousands if not millions of sensors. But it needs to be looked at in one single presentation, ideally in dashboard form, where a graphic presentation enables the best possible decision given the situation that is presented by the sum total of the collected data points. In the event of a threat assessment, how
does one structure facial recognition data and a criminal record, so that it can be linked to, say, a specific automobile, while tracking and understanding the meaning of that automobile’s vector on city streets? And how can all that data be seen in light of all the historical data collected by a “smart city,” while gathering up information from law enforcement, governmental sources, and even private sources? Putting aside for a moment the issues of policy and privacy, a predictive analytic capability will not truly exist until the data is properly structured, and stored so that it can be rapidly retrieved.
A PREDICTIVE ANALYTIC CAPABILITY WILL NOT TRULY EXIST UNTIL THE DATA IS PROPERLY STRUCTURED, AND STORED SO THAT IT CAN BE RAPIDLY RETRIEVED.
The policy that informs the use of big data is complicated, and it is exponentially more complicated when predictive analytics come into play, especially in the arenas of public safety. It is almost the stuff of science fiction when predictive analytic experts start to speak seriously about their ability to predict the likelihood of a crime occurring at a certain time and a certain location, or that a big data analysis could go back and retrieve the photo of a person wearing a certain color shirt on a certain street corner in a certain span of time that can be measured in microseconds. That said, a great deal of thought is already gone into the policies that have to be in place before big data can be collected and used in a meaningful, legal manner to serve people.

To look at just the example of police use of body cams, here are some of the policy issues that have to be addressed. A similar list of policy issues can easily be generated around the collection of data from public or private sources.

For police body cams, when a video is captured, it can fall into two general categories: transitory (everyday interactions) or evidentiary (video related to a crime). As video is captured, it needs to be relegated into one of these two general categories. But who decides? If it is the police alone, there will likely be public challenges to question the unilateral nature of that decision-making process. Next, how long is it kept? Even if video records fall into the transitory category, there has to be a determination as to how long the video is stored, because transitory data may become evidentiary in retrospect, if a crime is committed but only detected later. Next, who owns the data? For evidentiary video, whose property it is? Who has access to it? And under what conditions it is released? Is the video public property, very much like 911 calls are today? Can a news organization simply request the video, and put it right on the 5 o’clock news?
The questions just raised apply just to the storage of police video data, and those questions become all the more complicated and pertinent in the public policy arena when sophisticated predictive data analytic capabilities are brought to bear. For instance, should crime prediction based on publicly held data be allowed by freelancing private-sector companies? Or should the predictive analytic capability for crime prediction fall just to law enforcement? If the data is in the public domain, and predictive analytics prove to have efficacy in the world of crime, how far is law enforcement allowed to reach to collect data that might make those predictions even more accurate?
As with so many sectors where technology is advancing faster than the policy to control it, it is very likely that policy will be a trailing factor in predictive analytics rather than something that is put in place before capabilities develop.
5: VISUALIZING DATA
One of the lessons in the aftermath of the attacks of 9/11 is that various governmental institutions, in a fragmented way, possessed the information to very accurately determine who, when and where the attacks would occur. We just did not have the capability to synthesize it so that it was presented, collectively, in a single presentation format that enabled preventive action. Now, as big data capabilities are increasingly common, and algorithms to enable predictive analytics are also powerful as well, the question is how do we present the resulting intelligence in a way that enables a human to act. Many advanced practitioners of predictive analytics believe that the technology should automate all the machine to machine interactions, so that a human is presented with intelligence in a meaningful way, and allowed to use his or her intuition to make a decision. To present the results of a predictive analytic analysis in a meaningful way often means that the data should be displayed in a visual format, ideally in a dashboard presentation. For example, U.S. Customs and Immigration is potentially awash in large amounts of data about the travelers who were trying to enter the United States. As more and more information is available on an individualized basis for each traveler, and predictive analytics tools can use that data to determine if the traveler is a potential bad actor, then the ideal presentation of that data is on the screen in front of the customs agent who is deciding whether to issue a visa or not at an airport. If the traveler should not be admitted, the customs agent should see a red gumball. If the traveler should be admitted, the customs agent should see a green gumball. If the traveler needs to be interviewed further, the gumball needs to be yellow. The presentation of that green, red, or yellow gumball is really the very flower of the predictive analytics process. It seems almost childish that the result of all of that analytical power is the equivalent of a digital traffic light, but the solution is elegant in its simplicity.
The example cited above about providing a customs agent with a go/no-go/caution visualization instrument for deciding whether to admit a traveler to the US is an exemplary use case. Big data and predictive analytics are now being used to do everything from defend our nation from cyber attacks, two predict the likelihood of a traffic accident at a certain time of day on a certain corner.

**Medicine.** In the realm of personalized medicine, imagine if the capabilities of a sophisticated predictive analytics solution had access to unlimited patient records, with proper privacy protections in place. The predictive analytics could meaningfully lengthen the life of the patient, while reducing how often they are treated at the hospital, and driving down the rate of relapses and return visits to the hospital. It is very likely that patient records already exist to enable this kind of decision-making. The only challenge is structuring the data so it can be subjected to a predictive analytic analysis.

**Cyber Attacks.** Cyber attacks on the US are profligate today, and by the very nature they generate data at an astounding rate. The only thing missing to predict and prevent the attacks, in many cases, is the application of predictive analytics that have access to the preparatory steps that a bad actor must take to stage the cyber attack. Here too, the data exists, but it has not been meaningfully subjected to a predictive analytic tool, until very recently.

**Retail.** Predictive analytics also shows great promise in optimizing and fine-tuning the retail environment, so that shopping can be personalized, or that retail stores can be designed and staged in such a way as to optimize the experience, well enabling higher spends by shoppers. If predictive analytics has access to the customer profile, their purchase history, there personal history, and even such things as the time of day they like to shop (driven by facial recognition), the predictive analytical capability can help create an idealized shopping experience, and the retailer is pleased with how they were able to yield manage the spend of that shopper at an individualized level.
WHY ViON?
ViON offers a host of data analytics solutions that enable our customers to discover the value in their data, using those insights to form predictive patterns and trends. The DataAdapt platform of big data solutions can help uncover criminal activity, detect fraud, identify and prevent cyber threats or enable our customers to find the insights to improve the outcomes of their specific missions. With over 35 years of experience, ViON understands the power of having the right information in the right hands at the right time. To see how we can help you uncover the hidden insights in your data, visit us at www.vion.com.

ViON Headquarters
196 Van Buren Street
Herndon, Virginia 20170
(877) 857-ViON (8466)

www.vion.com