Predictive Analysis

Asiya Jaleel

ABSTRACT:

Predictive Analysis generally deals with extraction of data. It exploits pattern to find the historical transaction data to identify the risk. Use of Information Technologies allows automization of process for extracting of data that help to get interesting Knowledge and regularities, which means elimination of manual task and easier extraction of data directly from electronic records which will save lives an reduce the cost of health care services. Data Mining can enable health organizations to predict trends in paitent condition and their behavior which is accomplished by data analysis from different perspectives and discovering connections and relations from seemingly unrelated information. Healthcare data mining provides countless possibilities for hidden pattern investigation from these data sets. These patterns can be used by physicians to determine diagnoses, prognoses and treatments for patients in healthcare organizations.

Keyword: Data Mining, Health care, Data Collection, Enterprise data Warehousing (EDW).

I. INTRODUCTION

Predictive Analysis is a broad term describing about a variety of statistical and analytical technique used to develop models that develop models that predicts future events or behavior. It encompasses of a variety of statistical techniques from modeling, Machine Learning Data Mining that analyze current and historical facts to make future Predictions. Most of the models generate a score with a higher score indicating a higher likelihood of the given behavior or the event occurring. Data Mining is the component of the Predictive Analysis that entails analysis of data to identify trends, patterns or relationship among data. Predictive analysis exploits pattern to find the historical transaction data to identify risk. Predictive analysis is used in

a) Marketing.
b) Financial services.
c) Telecommunication.
d) Travel.
e) Health Care.

One of the most important application is Credit scoring where it deals with customer credit history, Loan application , Customer data in order to rank making future where credit payments is on time.
Ex: Credit card fraud occurrence, Fico Score.
It is also a term that comes from data mining which means extracting information from data and predict behavior pattern. Predictive analysis and data mining technique is much more efficient in business and marketing.
It offers the benefits

1. Enhance Business to improve decision making.
2. Manage the current challenges and boost the probability for future success.

Generally it is a business intelligence technology which produces the score for customer or an organization. In marketing predictive analysis helps managers to bring in more sales while spending less.

II. LITERATURE SURVEY

Predictive Analysis literature has three major content types: data collection & aggregation, model formation & production and delivery modalities. The literature suggests these three themes are the start, middle and end of a predictive analytic model's lifecycle. This paper will cover the data collection & aggregation and delivery modalities. Model formation and production methods have been excluded. There is enough research to write a paper on that alone as model formation relies heavily upon statistics methods such as logistic regression and neural networks. It is worth noting that since the paper covers the lifecycle of a predictive model and the topics overlap, some model production themes are found in the researched literature. Its' intertwined nature makes total exclusion impossible as related themes would be incompletely presented. The examination of the research focuses first upon the birth place of a predictive analytic model: the successful data collection methodology and how to best organize it.

Data Collection & Aggregation

The method with which data is collected, organized, used and stored makes it vital to the success of a project using predictive analytics. Included within this theme is how the data is

gathered, optimizing the data for use and finally the creation of a physical habitat where the data exists. All of these themes fall under the subject of data mining. Data mining could be studied alone or in non-predictive analytic applications, but here we examine data mining in the context of how it is successfully used to support predictive analytics.

Data collection

The first step in the data mining process is the collection of the data. This is the process of creating data variables representative of the physical world. Even before collecting the data, thought should be given to the planning and proposed use of data. To optimize a model, the business need should be thought of carefully. Then the data should be collected according to that need, resulting in data ultimately more suitable to a specific organizational applicatio. Shmueli and Koppius argue that this can be done by always using a data collection instrument which is relative to the prediction context and most meaningful to it. Some collection efforts have used a wide variety of patient centered information such as Diagnosis Related-Group (DRG) codes and then later narrowed the variables based upon final business need. The key insight here is that data should be catalogued in the form most true to itself. The cataloguing should be done in a form most true to the business need, which is supported by authors speaking about different fields.

Constraining dimensions

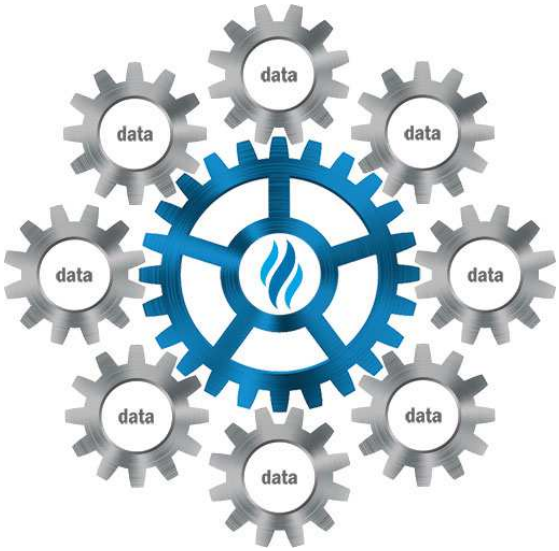
Once business need .and prediction contexts have been established, the collection of the data to be used in the model occurs. There are two keys in this process to success. One is that the data is constrained for the tailored model and the other is that the data is cleaned for high quality. Data constraints narrow down the amount of variables used. One method to constrain the data involves using the same constraints from many sites in a closed loop, thus making them scalable for an enterprise deployment elsewhere. Predictive accuracy will be improved upon such constraints as reducing the dimension of the prediction through data compression methods A reverse engineered method also exists where the model is trained to look for best examples from the data and then constrained from that. Once the dimensions are constrained, variants of other models can be created by an introduction of a new variable from the data set. The community here again suggests that the data be fitted to the business need. Using proper technique, the model can be strategically planned so that it can be applied in a local or enterprise basis dependent on business need. Literature suggests that the most successful dimensions of variables use the above methods to meet this need. Once the dimensions are in place, the data can be cleaned.

III. RESEARCH FOCUSED

There exists the seven reasons for Predictive Analysis in any organization.

a) Compete.
b) Grow.
c) Enforce.
d) Improve.
e) Satisfy.
f) Learn.
g) Act.

Start With an Integrated Data Warehouse and Analytics Measurement System



Enterprise Data Warehouse

We need data across the entire continuum of care in order to manage patient populations. This requires an enterprise data warehouse (EDW) platform. An EDW is the central platform upon which you can build a scalable analytics approach to systematically integrate and make sense of the data. Health Catalyst deploys a unique Late-Binding Data Warehouse that enables healthcare organizations to automate extraction, aggregation and integration of clinical, financial, administrative, patient experience and other relevant data and apply advanced analytics to organize and measure clinical, patient safety, cost and patient satisfaction processes and outcomes.

Foundational Dashboard and Reporting Applications

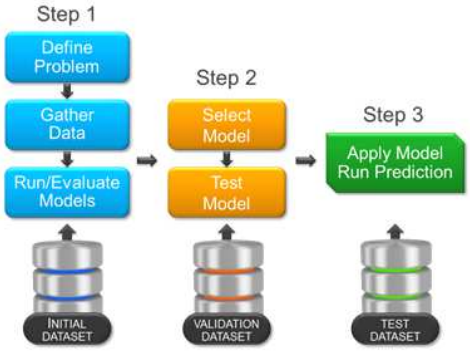
Today, most of your analytic resources are tied up doing manual reporting. New requests often take weeks or even months to clear the queue. You need to free up these analytic resources to focus on identifying and solving real problems instead of producing reports.

Health Catalyst provides a series of Foundational Applications which significantly accelerate the manual reporting work of many analysts into near real-time executive and department reports and dashboards. For example, Population Explorer, one of our most popular apps can produce the equivalent of 1000 dashboards with over 100 metrics.

Discovery Applications

Making sense of your volumes of population health data can be overwhelming. You need a systematic way to understand and evaluate the important opportunities and allocate resources to work on those key processes. Health Catalyst provides a series of Discovery Applications that will help you prioritize the opportunities within your system and help you focus on the areas of greatest opportunities in terms of quality of care improvement, cost reduction, revenue enhancement or patient injury prevention.Once you have your foundational enterprise data warehouse and analytics applications in place, you are better prepared to begin more advanced analytic tasks such as predictive analytics in the higher level of analytics adoption.

2. Use the Three Basic Steps of Predictive Modeling



The following is a simple schematic of the predictive modeling process. For predictive analytics to be effective, Lean practitioners must truly "live the process" to best understand the type of data, the actual workflow, the target audience and what action will be prompted by knowing the prediction.

1. The first step is to carefully define the problem you want to address, then gather the initial data necessary and evaluate several different algorithm approaches.
2. Step two refines this process by selecting one of the best performing models and testing with a separate data set to validate the approach.
3. The final step is to run the model in a real world setting.

The more specific term is prescriptive analytics, which includes evidence, recommendations and actions for each predicted category or outcome. Specifically, prediction should link carefully to clinical priorities and measurable events such as cost effectiveness, clinical protocols or patient outcomes. Finally, these predictor-intervention sets are best evaluated within that same data warehouse environment.

So many options exist when it comes to developing predictive algorithms or stratifying patient risk. This presents a daunting challenge to health care personnel tasked with sorting through all the buzzword and marketing noise. Healthcare providers need to partner with groups that have a keen understanding of the leading academic and commercial tools, and the expertise to develop appropriate prediction models.

IV. IMPLEMENTATION

Follow 4 Key Lessons Learned for Adopting Predictive Analytics in Healthcare Given that predictive analytics are listed as level 7 out of the 8 possible levels on the Healthcare Analytics Adoption Model, there are many keys and pitfalls that can occur at such a level if not properly prepared. Fortunately for healthcare, there are numerous existing models from other industries that can be combined with past healthcare examples to ease some of the potential pains and pitfalls. Highlights of some those key lessons include:

1. Don't confuse more data with more insight: While many solid scientific findings may be interesting, they do little to significantly improve current clinical outcomes.
2. Don't confuse insight with value: While many solid scientific findings may be interesting, they do little to significantly improve current clinical outcomes.

3. Don't overestimate the ability to interpret the data: Sometimes even the best data may afford only limited insight into clinical health outcomes.
4. Don't underestimate the challenge of implementation: Leveraging large data sets successfully requires a hospital system to be prepared to embrace new methodologies; this, however, may require a significant investment of time and capital and alignment of economic interests.

Health Catalyst Predictive Analytics Solutions

Health Catalyst not only has the expertise to develop prediction models, but our underlying warehouse platform is key to gathering the rich data sets necessary for training and implementing predictors. Notably, our prediction is only used "in context" – meaning when and where needed, with clinical leaders that have the willingness to act on appropriate intervention measures. Most important, however, these predictor-intervention sets can best be monitored and measured within that same data warehouse environment where otherwise not possible. Within Health Catalyst, data modelling and algorithm development is performed using industry leading tools for data mining and supervised machine learning such as Weka, Orange, and R. Ongoing efforts include classification models for a generalized predictor of hospital readmissions, heart failure, length of stay and clustering of patient outcomes to historical cohorts at time of admit. Most importantly, we have internal access to millions of de-identified hospital records in both the inpatient and outpatient settings and adult and paediatric populations. This training data is crucial to addressing the predictive analytics demands of clients and site customization.

V. CONCLUSION

Data mining has great importance for area of medicine, and it represents comprehensive process that demands thorough understanding of needs of the healthcare organizations. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve success of healthcare organization and health of the patients. Data mining requires appropriate technology and analytical techniques, as well as systems for reporting and tracking which can enable measuring of results.

VI. FUTURE

The Analytical techniques helps in resolving the risk which has occurred in the past based on several learning models .Based on the Expected model the future is correctly predicted.

REFERENCES

[1] Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases.American Association for Artificial Intelligence , 37-54.

[2] Candelieri, A., Dolce, G., Riganello, F., & Sannita, W. G. (2011). Data Mining in Neurology. In Knowledge-Oriented Applications in Data Mining (pp. 261-276). InTech.

[3] Bushinak, H., AbdelGaber, S., & AlSharif, F. K. (2011). Recognizing The Electronic Medical Record Data From Unstructured Medical Data Using Visual Text Mining TechniquesProf. Hussain Bushinak. (IJCSIS) International
Journal of Computer Science and Information Security, Vol. 9, No. 6 , 25-35.

[4] Eapen, A. G. (2004). Application of Data mining in Medical Applications. Ontario, Canada, 2004: University of Waterloo.

[5] Milovic, B. (2011). Usage of Data Mining in Making Business Decision. YU Info 2012 & ICIST 2012, (pp. 153-157).

[6] boirefillergroup.com. (2010). Data Mining Methodology. Retrieved 06 12, 2012, from Boire-Filler Group: http://www.boirefillergroup.com/methodology.php

[7] Zaïane, O. R. (1999). Principles of Knowledge Discovery in Databases. Department of Computing Science, University of Alberta.

[8] Milovic, B. (2012). Kompleksnost implementacije koncepta CRM. INFOTEH-JAHORINA Vol. 11, (pp. 528-532).

[9] Koh, H. C., & Tan, G. (2005). Data Mining Applications in Healthcare. Journal of Healthcare Information Management -Vol. 19, No. 2 , 64-72.

[10] Hosseinkhah, F., Ashktorab, H., Veen, R., & Ojaboni, M. O. (2009). Challenges in Data Mining on Medical Databases.