Using Data Analytics in a Hospital Setting

Lucia Croy

MIS530 – Predictive Analytics

Colorado State University – Global Campus

Dr. Edward Santiago

February 10, 2019

**Executive Summary**

The healthcare industry has been slow to adopt data analytics compared to many other industries due to a long list of reasons. Although there has been some progress in implementing data analytics systems there is still plenty of work to be done. One of the hesitations by hospital executives is ensuring that the organization does not get lost in the immense amount of data available and that the organization is able to efficiently use the data to improve processes. Wills (2014) discussed that the average amount of data a hospital would produce in 2015 would be 665 terabytes of data with it continuing to increase each year.

The hospital facing the issue of making sure that they have enough supply of specific medicines during peak virus season. Over the last few years the hospital has seen an influx of patients with a common virus during a similar time-period each fall. Using data analytics to predict the number of patients for the upcoming virus season will allow the hospital to ensure they have enough medicine on hand to treat all patients. This would allow them to account for a potential increase in patients from the previous year and not deplete their supply.

The recommended solution using predictive analytics would be to use a decision tree and a regression model focusing on seasonal variations. This would allow to identify patterns that occur during the same seasonal period each year. The algorithm would want to identify the expected number of patients coming in with the common virus and what medicine they were prescribed. The use of historical data could also be used to determine what the expected increase in patients each year and calculate that into the expected demand of the medicine. The data life-cycle would include a 5-10 years span in order to identify the pattern during the common 90-day period before expanding to look at the full calendar year.

SAS enterprise miner would be a good tool to help implement the analytics system. It allows the end-user to load the data and define what variable will need to be used in order to get the desired outcome. Upon initially building the regression model diagram it can be saved and built upon as needed should the hospital want to look at other information from the data set.

Overall the use of data analytics within the hospital will help improve the overall quality of care by ensuring that medicines are appropriately stocked to meet the needs of patients. As the hospital becomes more familiar with data analytics, they can then expand it to help improve other operational processes.

Using Data Analytics in a Hospital Setting

Data analytics is becoming a necessity in every industry in order to keep up with the changes in technology and demands of the market. Many large retail organizations use data analytics to develop marketing strategies and meet product demands from consumers. The healthcare industry however has been one of the last industries to start adapting data analytics into their processes. Many of the reasons for this is due to privacy concerns however the amount of data that is gathered in any medical organization or hospital could be used in a variety of ways to help improve processes and the overall success of the organization. This paper will review the use of data analytics in a hospital setting and discuss how data analytics can help resolve medicine constraints during the peak virus season which occurs in the fall. An introduction to the healthcare industry will be provided followed by look at the specific data problem, a suggested solution using predictive analytics, the data life cycle and how SAS Enterprise Miner can help sort through and understand the data.

**Introduction to the Healthcare Industry**

The amount of data available in the healthcare industry is continually increasing, however the industry has not implemented the data management systems needed to manage the data. Wills (2014) discussed by 2015 the average amount of data a hospital would produce would be 665 terabytes of data. With the estimated amount of data being 665 terabytes in 2015 one can only imagine that number has increased immensely now in 2019. Many healthcare organizations are still struggling to determine how data management solutions can help their organization and decide what data would be the most helpful to analyze. Many hospital executives are often overwhelmed by the amount of data available to them and fear that without the proper management it will only cause more problems than it will help the organization (Wills, 2014, p255).

The implementation of data management systems has met resistance due to several reasons. Wills (2014) listed some of those reasons being a provider’s desire for independence, inadequate technological infrastructure and disconnected systems (p. 255). Kaiser Permanente is one of the leaders when it comes to data analytics in the healthcare industry. Kaiser Permanente implemented a system called Health Connect which has saved approximately $1 billion dollars by allowing information to be shared within the organization allowing for the improvement of disease management (Wills, 2014, p 256). Privacy concerns has also been a concern within the healthcare industry in terms of data analytics regarding all the regulations and laws protecting patient information. Sharma, Panwar, & Sugandh (2018) discussed how providing security for patient data is the number one priority for healthcare organizations. This is especially due to the high-profile breaches and hackings that have been made public in recent years. Due to the immense amount and variety of data collected within healthcare organizations it is subject to an infinite array of vulnerabilities.

Sharma, et al (2018) discussed how there are a variety of things that can be tracked by managing and analyzing healthcare data. In relation to the data problem being discussed in this paper tracking what prescriptions are being giving to patients would help track and monitor the inventory levels of each medicine with in the hospital. The primary focus within each healthcare organization is improving patient care with the efficient handling of resources within each healthcare organization it would result in cutting costs and improving overall patient care (Sharma, et al. 2018).

**Data Problem**

The hospital is experiencing medicine shortage during the fall season due to an increase in patients with a common virus problem. Currently there is no system in place to keep track of the number of patients seen for the virus and prescribed the same or similar medicine which would allow them to predict the demand for the prescribed medicine during that time. Being able to monitor the demands for the specific medication during this time would allow for better planning to ensure more medicine is on hand. A hospital is like any other organization had can use data analytics to improve their operational processes which include inventory management. Loshin (2013) discussed how improving business processes permeates into all aspects of standard of care within a healthcare organization. The standard of care within hospital includes having enough medicine to provide to patients in need.

In terms of a hospital improving clinical operations could mean a variety of things one of those as mentioned above is managing the inventory of medicine kept on hand. Sharma, et al (2018) discussed how data analytics can support initiative to improve clinical operations within a healthcare organization by creating models and looking for patterns of various scenarios in order to proactively prevent specific events from occurring. In this case it would be the running out of the medicine to treat the virus that peaks during the fall season. By implement a data analytics system to help predict the amount of medicine needed during the fall virus season the hospital will not only be able to improve the level of patient care but also increase the number of patients seen due to keeping enough medicine in stock and not having to recommend that the patients go elsewhere.

**Solution Using Predictive Analytics**

Predictive analytics is generally defined as the use or application of mathematical algorithms on large data sets to predict a future occurrence (Hiller, 2016). Predictive analytics can be used to determine several things within a healthcare organization, however the number one priority of most organizations it to use predictive analytics to improve overall patient care. By developing a system to aid in medicine inventory management the hospital would help achieve the improvement of patient care by always being able to treat patients that come to their facility.

There are a variety of ways the hospital could get the information needed to better manage the medicine inventory levels. Data collected with each patient is loaded into the systems which also identifies what the diagnosis was and what medicines were prescribed. Since it has been determined that the is an influx of a common virus during the fall data from that 3-month period could be observed to forecast the expected medicine needed to get through the next season. Historical data from years back could also be pulled to provide a better data sample and get a more accurate average. Depending on how often the hospital receives shipments of medicines the time line could be changed to pull data every two weeks if needing to order more frequently. Hiller (2016) discussed how cost delivery and efficient delivery of high-quality health care are common goals for predictive analytics in healthcare (p 278). Providing quality care is important to every hospital and healthcare organization. When the hospital is not able to provide the needed treatment, the patient is forced to go to another facility which results in loss of revenue and potential loss of future patients.

Maisel and Cokins (2014) discussed a set of seven principles that are necessary to consider when determining and developing a data analytics process. With the current problem identified running out of medicine during peak virus season relates closely to principle which the problem must demonstrate a strong cause-and-effect relationship. Currently the number of patients that visit the hospital during the fall are often diagnosed with the same virus needing the same prescription medication. The demand on this medication results in the hospital running out and not being able to service the patients that come in after the medicine stock has been depleted.

Maisel (2014) also discussed the need to not only develop a successful business analytics function but also stressed the need to integrate a performance management system to ensure the use and success of the system. The performance management systems include the communication strategy, alignment within the organization, empowerment of employees once they understand the reasons behind the new system, focus and efficiency towards meeting the targets and execution of the plan (Maisel, et al, 2014, p. 1491). Without a plan that ensure the successful use and execution of the data system it is likely the system will fail, and the investment put in to the development of the system will be lost.

In order to predict the amount of medicine needed the hospital could use a both the decision tree and regression model to predict the quantity of medicine needed based on the expected patient volume. The information that would be needed to predict this information would be the date of visit, diagnosis, medication prescribed, medication on hand, next expected shipment. A decision tree is a predictive analytics model that is used to classify a data point like a flow chart (Cady, 2017, p. 101) It is called a decision tree because it takes the shape of a tree with data branching out based on what is defined in each node. The nodes in this decision tree could branch out with questions such as did the visit occur within the specified time frame, if yes then was the patient diagnosed with the fall virus, if yes was the patient prescribed the fall virus medication. The final branch to determine what the increase in patients was from one year to the next could be was the patient previously diagnosed with the fall virus. This would break down the data to show how many patients visited within the specified time frame, were diagnosed with the fall virus and prescribed the virus medication. The final question helps calculate the year over year increase of patients to calculate the average and add to future years medicine needs. The decision tree is a good starting point in terms of data analytics because it is easy for end users to understand especially those that may not have a background in data analytics (Cady, 2017, p. 102).

A regression model can be used to perform trend analysis which uses four major movements to characterize time series data (Han, Kamber & Pei, 2011, p 588). The four major movements include trend or long-term movements, cyclic movements, seasonal variations, and random movements. With this specific data problem seasonal variations would want to be the focus due to the common virus occurring the same time every year. Seasonal variation is described as nearly identical patterns that a time series appear to follow during the corresponding season over successive years (Han, et al, 2011, p. 588). This will help analyze the data for the specific time period in the fall to help determine the estimated number of patients expected to come in with the common virus and need the same prescribed medicine.

In the following sections more information will be discussed regarding the use of data analytics tools specifically SAS Enterprise Miner to create a decision tree and regression model to obtain the desired results. Prior to running the data through the data analytics tools all details of how the data is to be sorted and what the target variable is going to be. The forecasted amount of medicine needed should be 10-25% above what was needed the previous year in order to cover the potential increase in patients.

**Data Acquisition Life Cycle**

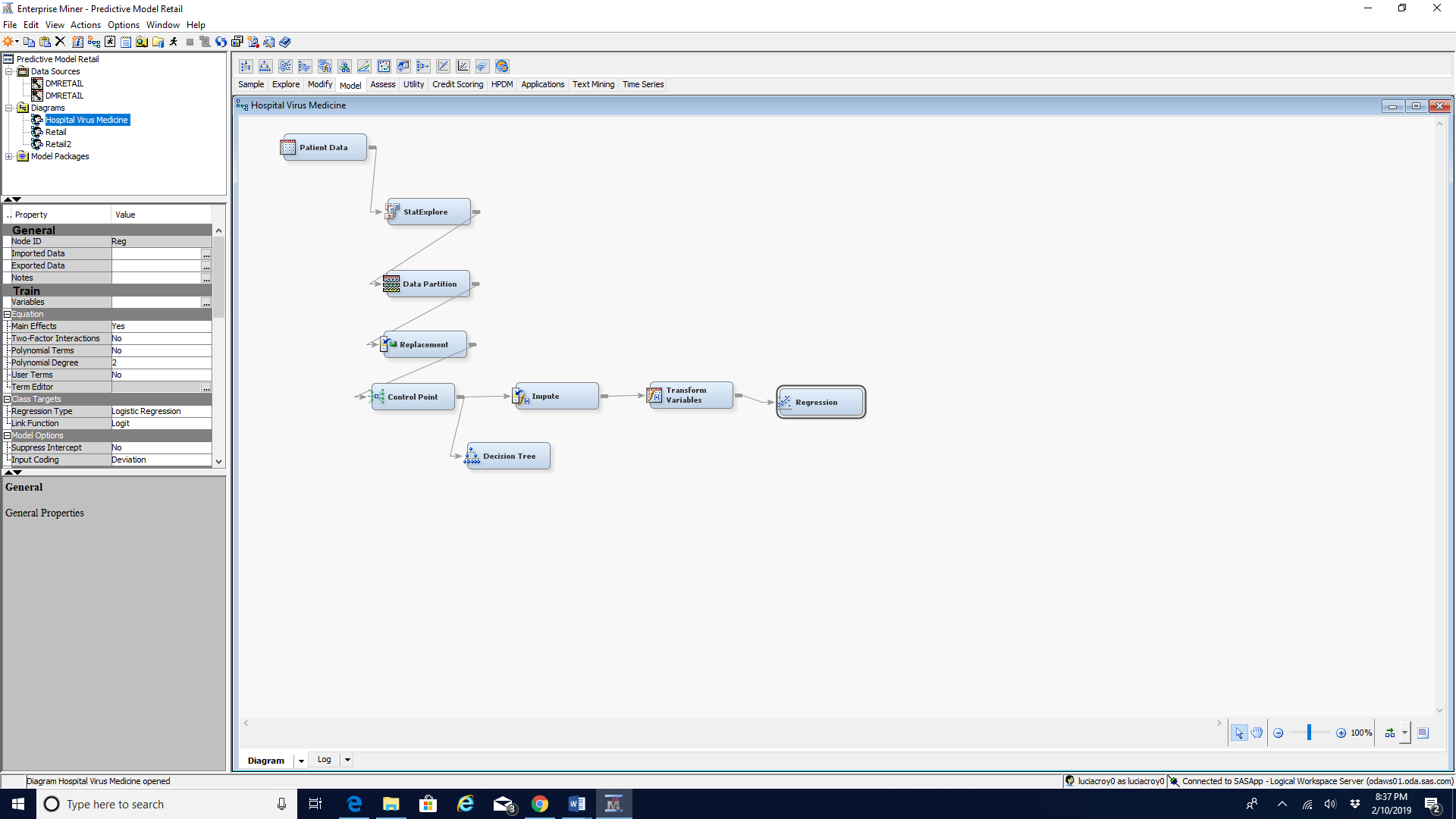
The data would need to be gathered continuously with historical data available at least 5-10 years. This would allow for better observation of relationships between symptoms and the patient diagnosis. Since this seems to occur during the same 90-day period in the fall the initial data could be run for that time period to determine the required amount of medicine needed to keep up with demand. Once the system has been implemented to predict the need for this initial time period the analysis could be expanded to include the entire calendar year to ensure that all patients no matter what time of year, they visit the hospital have access to the same medicine if needed. Specific data needed would be date of visit, diagnosis, prescribed medicine and whether the patient had been diagnosed with the fall virus previously and prescribed the virus medication. With all medical records being electronic all the data should be available in the patient records except for past diagnosis if the patient was seen at another facility. If there is concern for patient confidentiality the data can be scrubbed of any personal information in order to limit the use of and access to personal patient information.

**SAS Enterprise Miner**

SAS Enterprise miner is a great tool for organizations to use to make data analysis more efficient and user friendly. SAS enterprise miner could help in building a decision tree and regression model and defining what variables are going to be used to forecast the quantity of medicine needed for the upcoming virus season. As the diagram is built other models can also be used to determine if there is another type of model that may provide more accurate results such as decision trees and neural networks.

In order to use SAS Enterprise Miner a data source would need to be loaded containing patient information containing dates of visit, diagnosis and medication prescribed. The data would be processed through the STAT explore node, data partition, and replacement node to organize and prepare the data to run in the decision tree model. The Stat explore will show the standard deviation of the variables in order to help determine which variable will have the biggest impact on the target variable. The data partition node will identify what portion of the data is to be used for training and validation. The replacement node identifies what to replace missing variables with. In this case it would be important to sort out any data that is missing the variables of date visited and medication prescribed. The goal is to determine how much medication is prescribed during the fall virus season to proactively predict how much is needed to keep on hand and not run out. When identifying the properties for the decision trees it is important to specify the split rules to make sure that the data is sorted to get the correct results.

The regression model requires the addition of a couple additional nodes to clean the data even further. This enables the regression model to run off a smaller set of data removing variables that do not have an impact on the target variable the amount of prescribed medicine on hand. Figure 1.0 shows what the final diagram would look like to run the data provided in a decision tree and regression model.



*Figure 1.0*: Final diagram including decision tree and regression model. Captured from: SAS Enterprise Miner – Hospital Medicine Diagram

For further evaluation the results from both the decision tree and the regression model could be compared to determine which model provides the most accurate results based on the smallest standard deviation. Based on the historical data provided and run through the model a buffer can be added to ensure that there is always 10-25% more medicine on hand then what is needed in order to account for the potential influx of new patients as the historical data shows that the number of virus cases has continue to increase year over year.

**Conclusion**

Overall there is still several things that the hospital can do using business analytics, however, to start with implementing a system to manage medicine inventory will help improve their overall standard of care. Patients visit hospitals with the expectation of receiving treatment and medicine that will help cure or improve their symptoms regardless of what the visit is for. If the hospital is unable to provide treatment or administer the needed medications it results in lost revenue and potential patients due to the one patients experience. Patients look for recommendations when in search of what medical facilities to visit and ensuring that all patients receive the highest standard of care is the only way to receive referrals from existing patients and through reviews left online. The use of predictive analytics will assist in avoiding situations in which the hospital in not able to provide care to patients by allowing the hospital to always be prepared whether it be having enough medicine on hand, staff working, or available beds to meet the patient’s needs.

Edelstein (2013) discussed how data analytics will play a huge role in transforming and reforming the healthcare industry by generating positive results in improved patient care, reduced cost for both patients and hospitals and efficiency in resource utilization. Overall there is still many opportunities for the use of predictive analytics in the healthcare industry however to start with a smaller project like medicine inventory medicine can lead to the use in several other areas. It is becoming familiar with the systems and seeing the results that will get other departments and areas to use the system even if to improve simple things such as scheduling based on predicted volume.

References

Cady, F. (2017). The Data Science Handbook. Hoboken, New Jersey: Wiley Publishing.

Edelstein, P., M.D. (2013). Emerging directions in analytics. Health Management

Technology, 34(1), 16-17. Retrieved from

https://csuglobal.idm.oclc.org/login?url=https://search-proquest-

com.csuglobal.idm.oclc.org/docview/1282515951?accountid=38569

Getting Started with SAS Enterprise Miner: Building Decision Tree.

https://www.youtube.com/watch?v=IlUZYlgkeSc&list=PLVBcK\_IpFVi-

xzvJiOlf33UvVbRoLRu0z&index=3

Getting Started with SAS Enterprise Miner: Exploring Input Data and Replacing Missing

Values. https://www.youtube.com/watch?v=EBFtrPKQ0zg&list=PLVBcK\_IpFVi-

xzvJiOlf33UvVbRoLRu0z&index=2

Getting Started with SAS Enterprise Miner: Setting-up an Enterprise Miner Project.

https://www.youtube.com/watch?v=489wJm2X0TY&list=PLVBcK\_IpFVi-

xzvJiOlf33UvVbRoLRu0z

Getting Started with SAS Enterprise Miner: Compare Regression and Neural Network Model

Results.

https://www.youtube.com/watch?annotation\_id=annotation\_866713763&feature=iv&list

=PLVBcK\_IpFVi-

xzvJiOlf33UvVbRoLRu0z&src\_vid=IlUZYlgkeSc&v=TnWRJQb5z4c

Han, J., Kamber, M., & Pei, J (2011). Data mining: Concepts and techniques (3rd ed.). Waltham,

MA: Morgan Kaufmann Publishers.

Healthcare analytics & medical analytics market (predictive modeling, clinical analytics,

financial analytics & others) - trends & global forecasts to 2017. (2013, Nov 27). PR

Newswire Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-

proquest-com.csuglobal.idm.oclc.org/docview/1461927758?accountid=38569

Healthcare predictive analytics market to reach $19 billion by 2025 - growing efforts to bring

down healthcare expenditure - research and markets. (2016, Dec 21). PR

Newswire Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-

proquest-com.csuglobal.idm.oclc.org/docview/1850740977?accountid=38569

Hiller, J. S. (2016). Healthy Predictions? Questions for Data Analytics in Health Care.

*American Business Law Journal,* 56(2), 251-314. Retrieved from

https://doi-org.csuglobal.idm.oclc.org/10.111/ablj.12078.

Loshin, D. (2013). Business Intelligence: The Savvy Manager’s Guide. Waltham, Massachusetts.

Morgan Kaufmann.

Maisel, L. & Cokins, G. (2014). Predictive business analytics: forward looking capabilities to

improve business performance. Hoboken, New Jersey: Wiley Publishing.

Sharma, N., Panwar, A., & Sugandh, U. (2018). Big Data Analytics in Health Care: A Literature

Survey. *International Journal of Recent Research Aspects*, 5(1), 127-132. Retrieved from

http://search.ebscohost.com.csuglobal.idm.oclc.org/login.aspx?direct=true&db=aph&AN

=129311345&site=ehost-live

Wills, M. J. (2014). Decisions Through Data: Analytics in Healthcare. *Journal of Healthcare*

*Management,* 59(4), 254-262. https://doi.org.csuglobal.idm.oclc.org/10.1097/00115514-

201407000-00005