Applications in Predictive Analytics

Prediction As A Way to Better Health

Gurjus Singh

Northwestern University

Introduction

Healthcare has evolved quickly through technology in the past decade. New technological advancements, including data analytics, has allowed for a breakthrough in the treatment of illnesses such as cancer and has fundamentally shaped the way technology is used in healthcare. Analytics allows for making digital copies of medical records, finding new drugs, and discovering new genetic diseases (Brobriakav, 2018). One way in which analytics can be helpful is through prediction of the onset of diabetes. Diabetes is when the blood sugar has increased abnormally (NIH, 2016). One way this happens is when the pancreas, which is the source for insulin, cannot produce or react to this type of hormone well, known as Type 2 diabetes, most common in adults (NIH, 2016). Approximately 381.8 million people in the world have diabetes, and 42.8% of adults have not been tested for it (Nnamoko, Hussain, & England, 2018, paragraph 2). Through the use of Electronic Health Records, physicians have access to patient data to use predictive analytics. This type of system contains history of patient data such as doctor notes, treatment plans, and medicines patients are taking (Bhatia & Syal, 2018). With the implementation of an Electronic Health Records system and the use of key data science methods, physicians and personnel in the medical field can find patients with diabetes with better accuracy. Predictive methods that have been used to diagnose diabetes include clustering, data imputation, and ensembled supervised learning.

Methods

One method described in Bhatia and Syal (2018) was clustering such as K Means and K* Means . According to the paper, the K Means clustering algorithm was used on data provided from UC Irvine, whose population were from the Pima Indians female patients group (Bhatia &

Syal, 2018, para. 14). This data had attributes that were relevant such as body mass index, blood pressure and glucose concentration (Bhatia & Syal, 2018, para. 14). In using these two methods the prediction accuracy was between 95%-97% (Bhatia & Syal, 2018, para. 14). The purpose of using K-Means in the beginning of the study is to reduce the number of irrelevant attributes in the data (Bhatia & Syal, 2018). K Means employs input K and the number of clusters in trying to update data into those K Clusters (Garbade, 2018). K* Means differs from K Means because it involves more ways to effectively use the data provided (Bhatia & Syal, 2018). K*Means uses average feature values in the clusters to provide information of the features in that cluster in order to speed up the prediction process (Bhatia & Syal, 2018).

Another method that can be used is data imputation. With this method, physicians can use risk prediction models better (Masconi et al. 2015, para. 1). Imputation can be understood as a process in which missing data is replaced in different ways; for example, taking the average of similar features that are not missing (Masconi et al. 2015; Grace-Martin, 2019). Masconi et al. (2015) found about 30% of the patients that had missing data in the group (paragraph 16). According to the results, simple imputation method had better results indicated by the concordance statistic score, a number calculated used to indicate performance, in comparison to using the multiple imputation method; although, the purpose of the study was to persuade the reader that missing data resulted in lower prediction performance (Masconi et al. 2015; Austin & Steyerberg, 2012).

Lastly, another method that can be used is Ensembled Supervised Machine Learning (Nnamoko, Hussain & England, 2018). Ensembled Supervised Machine Learning is used to combine algorithms to get a better predictive model to cure diabetes (Nnamoko, Hussain & England, 2018). These combined algorithms first predict one-by-one and their results are combined to get an overall prediction score (Nnamoko, Hussain & England, 2018). According to the results provided, there are cases in which a combination of methods can be better than non-ensemble models; for example, in the study, the authors used three algorithms which performed better than using them individually or using more than three algorithms (Nnamoko, Hussain & England, 2018). With these useful methods available, many providers should reap the benefits of prediction of diabetes.

Management

The way in which management has changed through data and diabetes diagnosis is they have found better ways to prevent diabetes (Jayanthi, Babu & Rao, 2017, para. 4). For example, the way in which predictive models using data have been helpful is it finds which types of people will be more likely to get diabetes. Another way it can help is it can help predict physicians and other medical personnel predict blood glucose in ahead of time (Jayanthi, Babu & Rao, 2017). It can also help to predict the likelihood of a patient getting diabetes within five years or less (Jayanthi, Babu & Rao, 2017). Data in diabetes research can help physicians recommend diet changes, advise specific medication, and help patients come up with an adequate exercise regimen.

Conclusion

Medical research in the topic of diabetes is still continuing with profound impact on patients. With the help of data methods such as clustering patients, detecting missing data and combining methods together to come up with one prediction, there can be advancements in this diabetes research while helping timing of diabetes in patients slow down. Data is the way to go in the new age of computers and has shown it can be helpful in healthcare.

References

Austin, P. C., & Steyeberg, E. W. (2012, June 20). Interpreting the concordance statistic of a Logistic Regression model: relation to the variance and odds ratio of a continuous explanatory variable. Retrieved January 19, 2020, from https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-82

Bhatia, K., & Syal, R. (2018, May 14). Predictive analysis using hybrid clustering in diabetes diagnosis. Retrieved January 21, 2020, from https://ieeexplore-ieee-org.turing.library.northwestern.edu/document/8358313

Garbade, M. J. (2018, September 12). Understanding K-means clustering in machine learning. Retrieved January 18, 2020, from https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

Grace-Martin, K. (2019, February 26). Seven ways to make up data: common methods to imputing missing data. Retrieved January 19, 2020, from https://www.theanalysisfactor.com/seven-ways-to-make-up-data-common-methods-to-imputing-missing-data/

Masconi, K. L. (2015, September 25). Effects of different missing data imputation techniques on the performance of undiagnosed diabetes risk prediction models in a mixed-ancestry population of South Africa. Retrieved January 18, 2020, from https://go-gale-com.turing.library.northwestern.edu/ps/i.do?p=AONE&u=northwestern&id=GALE|A429828238&v=2.1&it=r

National Institutes of Health (NIH) (2016, December 1). What is diabetes? Retrieved from https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes

Nnamoko, N. N., Hussain, A., & England, D. (2018, October 4). Predicting diabetes onset: an Ensemble Supervised Learning approach. Retrieved January 18, 2020, from https://ieeexplore-ieee-org.turing.library.northwestern.edu/document/8477663

Jayanthi, N., Babu, B. V., & Rao, N. S. (2017, August 23). Survey on clinical prediction models for diabetes prediction. Retrieved January 18, 2020, from https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0082-7