A Prescriptive Analytics Model for Early Detection of Diabetes Propensity

Rajendra Desai
Head External Programs, ISME

Sirichandana V V
2nd Year- PGDM, ISME

Abstract

Many Prescriptive analytics applications in healthcare use patient data to enable early detection of the propensity of a patient developing serious diseases like diabetes, stroke, cancer, and heart disease. This study aims to build an analytics model utilizing readily available data to explore the possibility of early detection of a person’s propensity to develop diabetes. Changes in parameters in the blood and urine test reports of patients were used as predictors in the model. While experienced doctors would see this trend immediately if these records are stored and made available at the time of consultation, a model can serve as an early warning system as the data is being entered or shared with the customer.

We utilized data with a diagnostic laboratory on patients on an annual health plan with the laboratory. The data consisted of 1200 records and tracked the following variables Age, Hemoglobin, Glucose (Fasting) - FBS, Glucose (PP-2 hours), Urea, Creatinine, Cholesterol Total, Triglycerides (TGL), HDL Cholesterol, LDL Cholesterol, VLDL Cholesterol, Total Cholesterol/HDL Ratio, LDL/HDL Ratio, over a period of 2 years. Some of these patients tested positive for diabetes at various points in their visits to the clinic. Our data consisted of 70% men, 30% women with members in the age group of 25 – 55. The reported ranges of values of the selected predictor variables did not have gender variations.

We created multiple models using standard analytics techniques K-NN and Naïve Bayes similar to (Sing, 2014). Naïve Bayes model was more useful as the probabilities of the person developing diabetes were generated by the model and could connect to the end user needs. The system we are creating is basically a warning system which helps in shortlisting records that have higher probability of developing diabetes for further assessment. The results from our initial models had a high precision on out of sample data with a false positive ratio of 12%. It will be useful when patient volumes are high to identify the ones that require further investigation immediately. These models are currently under testing on live data inside the organization.

Keywords: Diabetes Detection, Naïve Bayes, Healthcare, Analytics Model.

Introduction

Diabetes is a chronic condition associated with abnormally high levels of sugar (glucose) in the blood. Insulin produced by the pancreas lowers blood glucose. Insufficient production of insulin (either absolutely or relative to the body’s needs), production of defective insulin (which is uncommon), or the inability of cells to use insulin properly and efficiently leads to hyperglycemia and
diabetes. The two types of diabetes are referred to as type 1 and type 2. Former names for these conditions were insulin-dependent and non-insulin-dependent diabetes, or juvenile onset and adult-onset diabetes. Over time, diabetes can lead to blindness, kidney failure, and nerve damage. These types of damage are the result of damage to small vessels, referred to as microvascular disease. Diabetes is also an important factor in accelerating the hardening and narrowing of the arteries (atherosclerosis) leading to macrovascular disease. Like strokes, coronary heart disease, and other large blood vessel diseases.

According to an ASSOCHAM report (2014) brought out on the occasion of World Diabetes Day there are around 68 million Indians suffering from diabetes and it is estimated that by 2035 this figure will go up to 125 million. The report also ranked the major cities based on percentage of diabetics of the total population. Delhi-NCR ranks the highest (42.5%), followed by Mumbai (38.5%), Ahmedabad (36%), Bangalore (26.5%) Chennai (24.5%).

Riccardo Bellazziet et. al (2015) pointed out that the capability of predicting the propensity of a disease is certainly a major challenge of biomedical research and clinical medicine. This may be due to complex interactions and variations in the patients’ known risk factors. Prescriptive analytics models are useful in the extraction of patterns in these interaction/variations. The aim of this study was building an analytics model utilizing readily available data to explore the possibility of early detection of a person’s propensity to develop diabetes. Changes in parameters in the blood and urine test reports of patients were used as predictors in the model. While experienced Doctors would see this trend immediately if these records are stored and made available at the time of consultation, a model can serve as an early warning system as the data is being entered or shared with the customer.

**Literature Review**

Factors affecting diabetes propensity, onset and control have been studied by many researchers Huang Y et all (2007) focused on identifying significant factors influencing diabetes control by applying feature selection to a working patient management system. They applied the three complementary classification techniques (Naïve Bayes, IB1 and C4.5) to the data to predict how well the patients’ condition was controlled. They identified the five important factors that influence blood glucose control as age, diagnosis duration, need for insulin treatment, random blood glucose measurements, diet treatment. Using these five factors their models could achieve 95% predictive accuracy and 98% sensitivity.

Miyaki K et al (2002) made a study to estimate the usefulness of Prescriptive analytics algorithms for identifying the best risk predictors of diabetic. They applied the Classification and Regression Trees (CART) method to the data collected from 162 type 2 diabetes mellitus patients. They found that age (cutoff: 65.4 years) was the best predictor, and depending on the age, the second-best predictor was body weight (cutoff: 53.9kg) for the group above 65.4 or systolic blood pressure for the group below 65.4. Bellazzi R and Abu-Hanna A (2009) discussed the prominent use of Prescriptive analytics in the context of diabetes management. They used blood glucose reading data generated from the regular monitoring of diabetes mellitus patients and from patients in the hospital intensive care unit. This data is maintained electronically and readily available for researchers, clinical practitioners, physicians and health care decision makers.

Thirumal P.C. and Nagarajan N (2015) applied various Prescriptive analytics techniques to predict diabetes mellitus. The data was sourced from the UCI Repository of Machine Learning Databases and consisted of 768 patients’, nine numeric variables. All the patients were female and were at least 21 years old. The data were analyzed with different classifiers such as Naïve

---

Bayes, Decision Tree, k-Means, SVM and k Nearest Neighbor. The study concluded that diabetes is increasing among young and old age people. The experiments concluded that kNN provided lower accuracy than the other methods. Mythili T et all (2013) and Rajwant Kaur and Sukhpreet Kaur (2013) have used Analytics models to predict heart disease. Models such as Genetic Neural Network, KNN+Genetic Algorithm, SVM-Decision Trees-Logistics regression and Genetic Algorithm were used in the analysis for the prediction of heart disease. Neural Network was found to be the best method for these models.

Adrien Jamain and David J. Hand (2005) have analyzed the anomalies in the Naïve Bayes method. They critically examined Zarndt’s Naïve Bayes and Statlog’s Naïve Bayes. The study concluded that Statlog’s Naïve Bayes is the better method. Sadhana and Savitha Shetty (2014) used eight different attributes and have shown how the techniques available in Hive and R could be used in diabetes prediction. Riccardo Bellazzi, Fulvia Ferrazzi and Lucia Sacchi (2015) have discussed the essentials of application of Prescriptive analytics and data analytics technology in clinical medicine. Their study reviewed the features of predictive clinical Prescriptive analytics with special focus on the possibilities of using these methods to translate molecular medicine result into clinically useful Prescriptive analytics models.

Saidhana and Savitha Shetty (2014) have shown how the techniques available in Hive and R could be used in diabetes prediction. Riccardo Bellazzi, Fulvia Ferrazzi and Lucia Sacchi (2015) have discussed the essentials of application of Prescriptive analytics and data analytics technology in clinical medicine. Their study reviewed the features of predictive clinical Prescriptive analytics with special focus on the possibilities of using these methods to translate molecular medicine result into clinically useful Prescriptive analytics models.

Sigurdardottir A K et al (2007) analysed the factors that contribute to improvement in glycemic control in educational interventions in type 2 diabetes They applied the CART C4.5 using WEKA (Waikato Environment for Knowledge Analysis) data-mining software in order to identify the best factors that predict changes in glycated hemoglobin (HbA1c) level. The study concluded that the effect of the factor “diabetic education intervention” in diabetes treatment is significant and achieved a notable drop (.8–2.5%) in HbA1c levels. The other factors such as duration, education content, and intensity of education have no impact on changes in HbA1c. Wright A et all (2005) applied the reconstructability analysis, an information-theoretic Prescriptive analytics technique, on the MQIC data set to empirically identify risk factors for various complications of diabetes. The results found that the best predictor of microalbuminuria was an elevated urine microalbumin and reconstructability analysis is a sufficiently robust technique to discover valid associations in a large clinical data warehouse. Eleni I Georgaet all (2011) worked on glucose prediction in Type 1 and Type 2 diabetic patients using multivariate, nonlinear and dynamic interactions in glucose metabolism. Their study proposed four different models for the glucose prediction. The study concluded that the glucose concentration in type 1 diabetic patients can be predicted with a sufficient numerical accuracy.

Abdullah Aljumah and Mohammed Khubeb Siddiqui (2014) investigate the interrelationship between hypertension and diabetes risk factors using SVM based Prescriptive analytics technique. The study used the medical history record of patients on several aspects. The study found that Hypertension and Diabetes have a strong relationship and the risk factor diet had the highest prediction value as compare to other risk factors specially in the age group of more than 55 years.

Data and Methodology

We utilized data available with a diagnostic laboratory on patients on annual health plans with the laboratory. The data consisted of 487 records and tracked the following variables Age, Hemoglobin, Glucose (Fasting) - FBS, Glucose (PP-2 hours), Urea, Creatinine, Cholesterol Total, Triglycerides (TGL), HDL Cholesterol, LDL Cholesterol, VLDL Cholesterol, Total Cholesterol/HDL Ratio, LDL/HDL Ratio over a period of 2 years. Some of these patients tested positive for diabetes at various points in their visits to the clinic. Our data consisted of 70% men, 30% women with members in the age group of 25 – 55.
We evaluated multiple models using standard classification techniques K-NN and Naïve Bayes similar to (Sing, 2014). Naïve Bayes model was more useful as the probabilities of the person developing diabetes were generated by the model and could connect to the end user needs. Our model will be part of an early warning system to help shortlist records that have higher probability of developing diabetes for further assessment.

Managerial Implications for Diabetes Management

Early detection of diabetes propensity can result in prevention and cure of diabetes. This can be through lifestyle changes, dietary corrections and medication. If diabetes doesn’t get diagnosed and treated in time it might lead to complications like blindness, stroke, kidney failure, and stroke. Although pre-diabetes takes many years to progress, if left untreated it leads to the development of type 2 diabetes and increased risk of microvascular and macrovascular diseases and their complications.

This paper highlights the use of prescriptive analytics models for prediction of diabetes propensity using data on standard tests offered in health plans by diagnostic laboratories. An important contribution of this paper is highlighting the importance of utilising the ‘change in the point values’ of test results which produced a better prediction model than using the point values. To handle diabetes better, patients need to be aware of their condition early. Health Care organisations can build this model into their databases and provide the service to their customers.

References