ISSN No:-2456-2165

Demonstrating Predictive Analytics in a Sample of Coronary Artery Disease Patients

Dr (PT). Tania Mahapatra
Medico Research Analyst
Samapriya Health Data Analytics,
New Delhi, India

Abstract:- Predictive analysis is using the big data from health care for analyzing and forecasting an occurrence or an event. It is fast gaining popularity and opening many avenues for health care industry, especially assisting clinicians in decision making. The technology involves running the historical data through a suitable algorithm to create a predictive model which then can be applied on sample data. This paper illustrates using predictive analytics in cardiac patients. Coronary Artery Disease is one of the major causes of adult mortality that is due to narrowing of blood vessels supplying the heart and is associated with abnormal levels of fasting sugar, cholesterol, heart rate, blood pressure etc. The objective of the study is to analyze the physiological parameters through a predictive model and ascertain whether a sample population of cardiac patients are having lumen narrowing or not.

Keywords:- *Predictive Analytics, Health Data Analytics, Big Data, CAD.*

I. INTRODUCTION TO CAD

Human heart is a strong muscular organ that pumps approximately 3000 gallons of blood daily. But to work efficiently, its own arterial supply also needs to be continuous which is maintained by the coronary arteries. If there is a blockage to these arteries, cardiac muscles are deprived of sufficient oxygen and nutrition eventually giving rise to symptoms of Coronary Heart/Artery Disease. The pathology usually begins with injury to the innermost layer of the coronary artery that leads to deposition of fats, cholesterol and other cellular material at the site. Build-up of these fatty deposits or plaques cause physical obstruction to blood flow and further changes in the tonicity of lumen causing hardening of the vessels. This is called as atherosclerosis [1] which precipitates angina, nausea, dyspnea and other symptoms. If the clog is big enough, it causes coronary thrombosis that in turn causes cellular damage to the cardiac muscles and the resultant cardiac arrest. The modern lifestyle practices of smoking, drinking, sedentary work and increased fat consumption are the potentiating factors for CAD [2].

A recent study by Registrar General of India in partnership with health experts found that 68.4 percent of all cardiovascular deaths were due to coronary artery disease in the patients belonging to age group 30 to 69 years. Unhealthy lifestyle practices worsen the other risk factors like hypercholesterolemia, hypertriglyceridemia (>200 mg/dl), low HDL cholesterol (<40 mg/dl), hypertension (SBP >140), obesity and diabetes mellitus [3]. Some studies have also reported the impact of age and gender on prognosis, showing women above 50 years and men above 40 years more susceptible to CAD. Generally, men are more vulnerable to coronary disease probably due to increased job stress and responsibility of livelihood [4].

Usually CAD progresses over a period of time and patients tend to address their symptoms casually. The lumen of the arteries progressively decreases until a major cardiac event happens during instances like excessive physical exercise or a stressful life incident [5]. The patient experiences sweating, tightness or heaviness in the chest with either aching or burning pain that typically radiates to neck, jaw, shoulders, arms and back [6]. At other times, there may be silent signs, especially in diabetics, called atypical symptoms like fatigue, wrist pressure, indigestion, stomach pain or even snoring [7]

Treatment for CAD involves self-care, medications and surgical procedures. Lifestyle changes like quitting smoke and excessive fat in diet is suggested along with appropriate physical exercises. Pharmacological measures like statin, beta blockers, calcium-channel blockers, anti-coagulants, ACE inhibitors and ARBs form the major interventions [8]. However, in severe blockages, surgical procedures like CABG and Angioplasty with stent is required to restore blood flow to the heart muscles [9]. Delay in diagnosing will delay adequate medical care and increase the probability of mortality.

II. PREDICTIVE ANALYTICS

A. Introduction

Predictive analysis uses a mathematical model built on principles of statistics, data mining and machine learning. There are many algorithms available for predictive model and no one algorithm can be stated as best for all problems. The size and structure of dataset decides which statistical algorithm will be used to create a set of rules that in turn will govern how to predict the values of the target variable. Literature has shown that using predictive analytics has proved helpful in clinical decision making, forecasting patient compliance and scheduling optimal patient-doctor ratios [10] – [13].

B. Applications in health care

Predictive analysis can contribute in assessing a patient on deeper level and improve public health care. Digitization of health care industry has resulted in storage of patient data including their medical histories and health statistics. For cardiac patients, the important parameters of interest are heart rate, blood pressure, cholesterol levels, ECG recordings, presence of associated conditions like diabetes and likewise [14] – [16]. The biggest peril these cardiac patients face is narrowing of coronary arteries that can bring about angina and failure of cardiac muscles. To have a system that can factor in all the physiological parameters to forecast the risk of lumen narrowing, in order to provide timely medical or surgical intervention is of importance here. This technology, to use Big Data, for predicting certain occurrences or instances is called predictive health data analysis [17].

III. ILLUSTRATION OF PREDICTIVE ANALYTICS

A data of 303 heart patients was adopted from open-source University of California Irvin Machine Learning Repository at http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/ The database included 14 parameters, including demographic and physiological types, which included chest pain, heart rate, sugar levels to ECG changes at rest and exercise. The target parameter was whether lumen narrowing was 'present' or 'absent' in these patients (Fig. 1). This data was fed into a learning regression model available through analytics software.

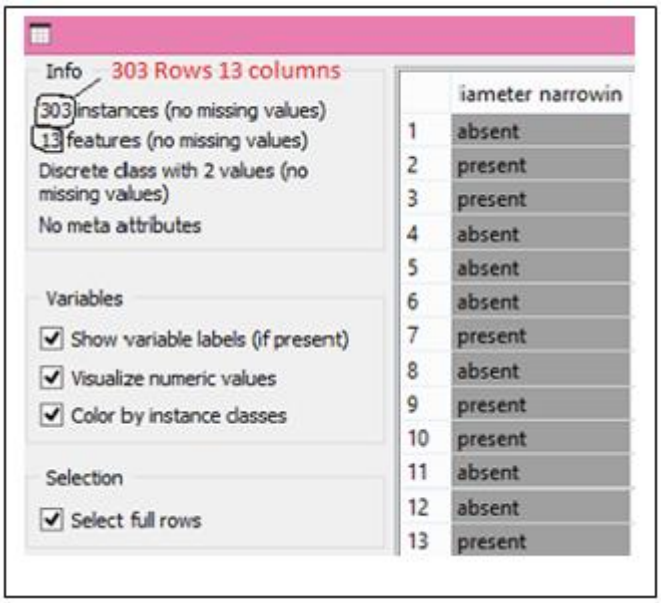

Fig 1:- Learning Data with Target Variable – Diameter Narrowing

Since our problem belongs to the binary classification (absent or present) type with independent variables affecting our target dependent variable (narrowing of lumen diameter), the most suitable algorithm will be logistic regression. It is also called logit regression as it predicts the probability of an instance happening through a logit function [18]. The logistic function is represented by an S shaped curve on a graph showing curvilinear relationship. The model created for our problem showed accuracy level of 85 percent which is considered a good indicator (Fig. 2).

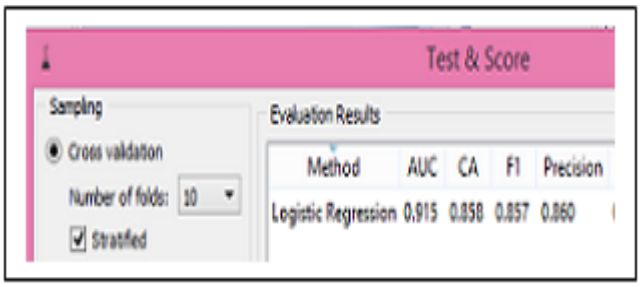

Fig 2:- Accuracy Level of the Selected Predictive Model

The logistic regression model was trained on 303 patients' data and was then made to use the learning model on a hypothetical sample data of 10 patients, to predict whether there might be lumen narrowing or not.

The result of our predictive model (Figure 3) showed that three out of ten patients were likely to have lumen narrowing in accordance to the historical data of 303 patients that was fed into the learning algorithm (Figure 4).

It is worth mentioning that the algorithm took into account the physiological parameters with respect to the given condition of the patients at the time. It does not mean that the other seven patients are free of CAD risk for ever. With changes in factors, for instance, age, cholesterol levels and fasting sugar level etc. there will change in data that would need to be run again through the system to know the current probability.

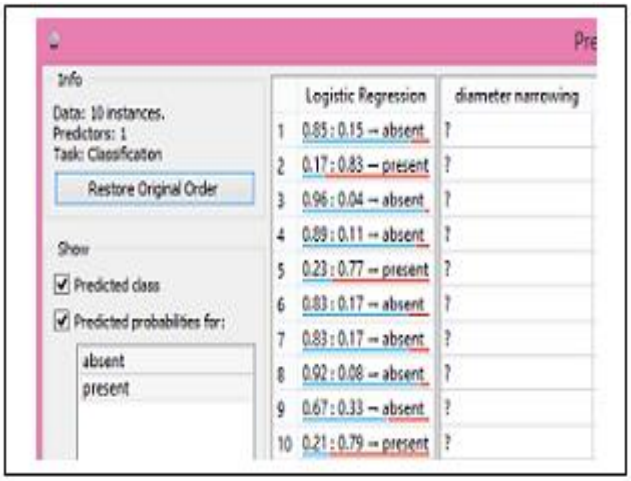

Fig 3:- Results of Predictive Analytics on 10 Sample Data

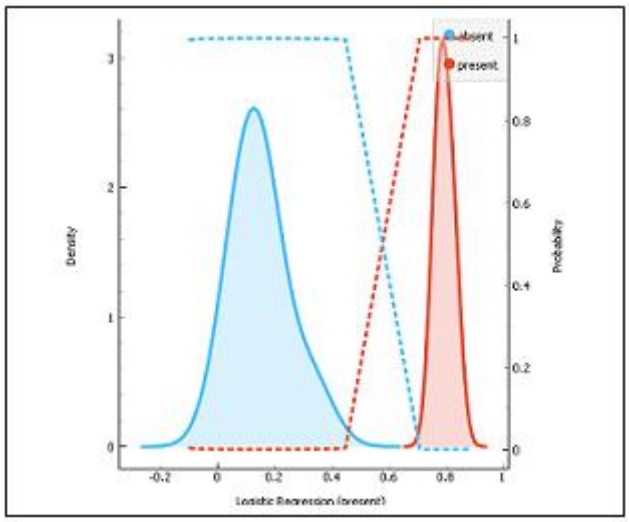

Fig 4:- Graph Representing the Logistic Regression Results

IV. CHALLENGES IN USING PREDICTIVE HEALTH ANALYTICS

A predictive model may recommend certain interventions which might not be in coherence with a physician's perspective [19]. This might include withholding a treatment because of significantly lower probability of an occurrence, for example, in our case study the patients who were excluded from having lumen narrowing do not predictably require costly angiography.

History of abuse and discrimination [20] may heighten concerns that Big Data can be employed in identifying vulnerable high-cost patients in society only to be neglected and discriminated further.

A predictive model runs in accordance to priority variables or factors that are fed into it. Hypothetically, a hospital might be more interested in patients having medical coverage, a doctor might be more focused in time management, the diagnostic center might be engrossed in income generation and a patient might be concerned with fewer visits and quick recovery. To remove bias, during model development, all concerned stakeholders must be taken into account.

Every predictive model should undergo rigorous evaluation and a check for reliability as there is substantial risk to patients' health in case of a wrong judgement. Painstaking validation of performance indicators and sensitivity of results should be done to improve the functionality of the model.

V. CONCLUSION

Predictive health analysis is a promising technology that has the potential to assist health care providers in clinical decision making. However, there should be a firm ethical foundation upon which the technology is delivered to avoid turning this boon into a societal curse.

REFERENCES

[1]. G. K. Hansson, "Inflammation, atherosclerosis, and coronary artery disease," New England Journal of Medicine, vol. 352, no. 16, pp. 1685–1695, Apr. 2005.
[2]. V. Fuster, L. Badimon, J. J. Badimon, and J. H. Chesebro, "The pathogenesis of coronary artery disease and the acute coronary syndromes," New England Journal of Medicine, vol. 326, no. 5, pp. 310–318, Jan. 1992.
[3]. D. Jain, "Heart disease deaths in India: What statistics show," https://www.livemint.com, 26-Jul-2018. [Online]. Available: https://www.livemint.com/Politics/fKmvnJ320JOkR7hX0lbdKN/Rural-India-surpasses-urban-in-heart-diseaserelated-deaths.html. [Accessed: 15-Apr-2019].
[4]. F. Sanchis-Gomar, C. Perez-Quilis, R. Leischik, and A. Lucia, "Epidemiology of coronary heart disease and acute coronary syndrome," Annals of Translational Medicine, vol. 4, no. 13, Jul. 2016.
[5]. N. O. R. A. Goldschlager, A. Selzer, and K. Cohn, "Treadmill Stress Tests as Indicators of Presence and Severity of Coronary Artery Disease," Annals of Internal Medicine, vol. 85, no. 3, pp. 277–286, 1976.
[6]. A. A. Chun and S. R. Mcgee, "Bedside diagnosis of coronary artery disease: A systematic review," The American Journal of Medicine, vol. 117, no. 5, pp. 334–343, Sep. 2004.
[7]. D. B. Pryor, F. E. Harrell Jr, K. L. Lee, R. M. Califf, and R. A. Rosati, "Estimating the likelihood of significant coronary artery disease," The American journal of medicine , vol. 75, no. 5, pp. 771–780, Nov. 1983.
[8]. R. Collins et al, "Blood pressure, stroke, and coronary heart disease: part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context," The Lancet, vol. 335, no. 8693, pp. 827–838, Apr. 1990.
[9]. N. K. Wenger, "Gender, coronary artery disease, and coronary bypass surgery," Annals of Internal Medicine, vol. 112, no. 8, pp. 557–558, Apr. 1990.
[10]. Y. Wang and N. Hajli, "Exploring the path to big data analytics success in healthcare," Journal of Business Research, vol. 70, pp. 287–299, Jan. 2017.
[11]. W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," Health information science and systems, vol. 2, no. 1, p. 3, Dec. 2014.
[12]. J. Sun and C. K. Reddy, "Big data analytics for healthcare". In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 1525-1525, August 2013.
[13]. Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," Technological Forecasting and Social Change, vol. 126, pp. 3–13, Jan. 2018.
[14]. R. Gianrossi et al, "Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-

analysis.," Circulation, vol. 80, no. 1, pp. 87–98, Jul. 1989.
[15]. B. F. Robinson, "Relation of Heart Rate and Systolic Blood Pressure to the Onset of Pain in Angina Pectoris," Circulation, vol. 35, no. 6, pp. 1073–1083, Jun. 1967.
[16]. M. Romano, "Postexercise systolic blood pressure in coronary artery disease," American Heart Journal, vol. 121, no. 6, pp. 1847–1848, 1991.
[17]. A. Kankanhalli, J. Hahn, S. Tan, and G. Gao, "Big data and analytics in healthcare: introduction to the special section," Information Systems Frontiers, vol. 18, no. 2, pp. 233–235, Apr. 2016.
[18]. I. G. Cohen, R. Amarasingham, A. Shah, B. Xie, and B. Lo, "The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care," Health Affairs, vol. 33, no. 7, pp. 1139–1147, Jul. 2014.
[19]. T. B. Murdoch and A. S. Detsky, "The Inevitable Application of Big Data to Health Care," Jama, vol. 309, no. 13, pp. 1351–1352, Apr. 2013.