Predictive Analytics on Healthcare: A Survey

S. Divya Meena[1], M. Revathi[2]

[1]ME-CSE, Kingston Engineering College, Vellore 632059, India

[2]Assistant Professor, Department of Computer Science, Kingston Engineering College, Vellore 632059, India

Abstract: ***Healthcare is indeed a considerable pointer for the development of society. Health does not only mean as dearth of disease but also capability to apprehend one's potential. In reality, there is a big gap between the rural and urban health service facility and accessibility. This paper identifies some of the problems in Indian healthcare and attempts to provide a solution by exploring the capabilities of healthcare. So, the services rendered by healthcare are not a mere responsibility of medical field but also of information technology. In fact, data mining plays an active role in providing a consistent accuracy in predicting the diseases and its risk factors. Some of the data mining applications and techniques used in real world are discussed.***

Keywords: Predictive Analytics**,** Health Management System, Insurance, Co morbidity Index, LO

1. Introduction

Hospitalization is the most prevalent component of health expenses. Healthcare administrators on their part are striving to lower the cost of care at the same time, improving the quality of care given. The rising cost of health care is one of the world's most important problems. Accordingly, predicting such costs with accuracy is a significant first step in addressing this problem. In addition, identifying and managing patients with high risk of survival is important not only for the individual but also for the government, hospital and health insurers.

The value of health insurance claims data in medical research has often been questioned because these databases are designed for financial reasons and not for clinical purposes. Nevertheless, claims data has been shown to be useful in many settings and is increasingly used for medical research. The use of health risk assessment methods based on medical diagnosis codes from administrative claim data continues to grow.

2. Data Mining

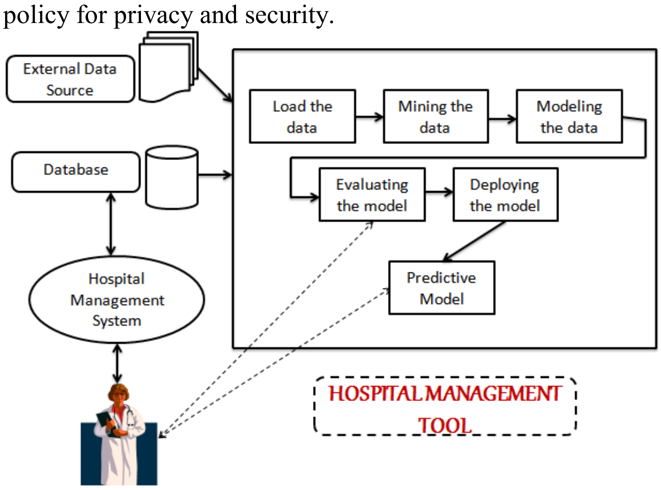
Data mining is defined as the analysis of often large observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Data mining in the field of Healthcare is a gradually developing methodology that is used to identify a legitimate, useful relationships and patterns in the data that are too complex and subtle for human to make out. The previously unknown patterns and relationships in a database can be easily identified with a Data mining algorithms like classification, association, clustering, etc and with this we can build a Predictive Model. To identify the unknown patterns, we use the life cycle model of data mining. It includes all the process from data selection to data exploitation.

2.1 Applications of Data mining in the field of Healthcare:

Data mining is beneficial for all those who are involved in the healthcare including Healthcare Insurers, Healthcare organizations, Doctors and patients. For healthcare insurers, data mining can help identify Fraud and Abuse in the insurance. For Healthcare organizations, it supports Hospital Management system, Patient Tracking system and thereby forming a strong Customer relationship management. For doctors and physicians, it can help predict the LOS, mortality rate, risk factors, and re-admission rate. With all the aforesaid, a doctor can choose upon the effective treatment available for the patient. And finally, a patient can get a quality treatment at affordable cost. According to a survey, 87% of the deaths in the US hospital have been prevented with the Predictive tool.

2.2 Advantages of Data Mining in Healthcare:

With the advent of Information Technology in the field of healthcare, the paper based prescription has turned to Electronic records, which is highly reliable. The electronic record is capable of storing voluminous data and this includes patient demographic, socio-economic data, previous medical data, laboratory values and much more. These data though private are essential for providing an effective treatment to a patient. Data mining helps in providing a clear policy for privacy and security.



Licensed Under Creative Commons Attribution CC BY

3. Predictive Analytics

One of the major challenges in the field of healthcare is the quality of service and affordable cost for the service rendered. Quality treatment should not backfire on the cost of the treatment. Predictive Analytics is actually a big Data Initiative in USA. With a predictive Analytics tool, Doctors will just be a consultant and Readmission of patients will be a problem for the hospital. So, the US health organizations are striving hard as not to get a patient re-admitted. The treatment turns from Pay-per-service system to Outcome – driven payment System. Proper diagnosis should be done and an appropriate treatment should be carried out. At the end, every single detail has to be recorded for further analysis. Computer-based information can server this purpose. HPN, Heritage Prize Network was conducted by US as a competition to improve healthcare by early prediction and a prize amount of $3,000,000 was awarded for the best algorithm. Predictive analytics enables more effective monitoring and greater cost-efficiency for a healthier world.

4. Charlson Co Morbidity Index

Charlson Co morbidity Index is a health based tool that is used to appraise the co morbidity risk of a patient, so that a medical specialist can make informed decision about the medical procedure to be carried out on the patient. The Charlson index takes into account of several health conditions along with the age factor. The health conditions are based on the ICD, International Classification of Disease diagnosis codes. The term Co morbidity refers to the effects or existence of one or more additional conditions that exist independently or dependent on the primary conditions. CCI, Charlson Co morbidity Index includes several categories of the co morbid conditions for predicting the mortality rate and risk factor. With this index, we can predict both short term and long term benefits of a treatment. For instance, if a patient is diagnosed with cancer as a primary disease but also has chronic heart failure and renal disease as co morbid conditions, then the short term benefit is no more important than the cost and risk of the treatment.

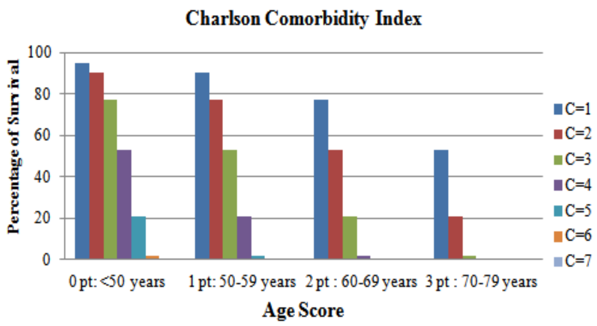
4.1 CCI construal

The Charlson Co morbidity Index is constructed with several diseases and a score is allotted for each disease based on it severity and mortality risk. The score is basically 1, 2, 3 or 6. More the points, more the adverse outcome is. The following table displays the Charlson Co morbidity Index.

Condition	*Points*
Myocardial Infarction	1
Congestive Heart Failure	1
Peripheral Vascular disease or bypass	1
Cerebrovascular disease	1
Hemiplegia	2
Pulmonary disease	1
Diabetes	1
Diabetes with end organ damage	2
Renal disease	2
Mild liver disease	2
Severe liver disease	3
Gastric or peptic ulcer	1
Cancer	2
Metastatic solid tumor	6
Dementia or Alzheimer's	1
Rheumatic	1
HIV or AIDS	6
Hypertension	1
Skin ulcer	2
Depression	1
Warfarin	1
Non-metastatic solid tumor	2
Malignant lymphoma	2
Leukemia	2

4.2 Interpretation of Charlson Co morbidity Index:

1. Calculate the Charlson Co morbidity Index value "i", by awarding the appropriate points to each of the condition you are diagnosed with.
2. Add the Age score to the value of "i"
   a. Age <50 years : 0 points
   b. Age 50-59 years : 1 points
   c. Age 60-69 years : 2 points
   d. Age 70-79 years : 3 points
3. The sum of value "i" and age score gives the value of "C" ie., C=Age score + i
4. Calculate the Charlson probability with Z=0.983^ (e ^(C*0.9)), where Z is the 10 year probability of survival.



5. Survey of Literature

Healthcare Management System is achieved with the advent of Predictive Analytics. While the baseline of a Predictive Analytic Tool is the same, the tool comes in various versions. Several developers aim at developing a tool that is efficient than all other tools developed so far. The following papers have implemented some of the most prevalent tools.

Licensed Under Creative Commons Attribution CC BY

Author	Technique Used	Rationale
Yilmaz Goksen et al	Estimation, Prediction, Classification, Clustering, Association	Enhancement Of Strategic Decisions In Medical Records Using Data Mining
Shweta Kharya et al	Classification, Neural Network, Association Rule Mining, C4.5 Decision Tree, Naïve Bayes.	Diagnosis And Prognosis Of Cancer Disease
Aqueel Ahmed et al	Association Rule, Clustering	Diagnosis And Prognosis Of Heart Disease
Taranath NL et al	Knowledge-Based And Learning Based System	Decision Support System For Medical Record Data
Abhishek Taneja et al	Neural Network, Decision Tree, Naïve Bayes	Prediction Model For Heart Disease
Matthew Herland et al	Big Data Tools	Role Of Big Data In Healthcare Using Data Mining
Gitanjali J et al	Apriori Algorithm	Medical Data Mining For Frequent Disease Identification
Jacques Donze et al	Striving For Quality Level And Analyzing Of Patient Expenses [Sqlape]	To Prevent 30-Day Hospital Re-Admission
Enrico Coiera et al	Charlson Comorbidity Index	To Predict The Cumulative Risk Of Death
Hude Quan et al	ICD-9-CM , ICD-10-CM, Charlson Comorbidity Index	To Update The Charlson Co Morbidity Index And Risk Score In Hospital Data
John Billings el Al	PAAR-30	To Predict The Risk Of In-Patients Getting Readmitted Within 30-Days Of Discharge.
Naren Meadem et al	Logistic Regression, Naive Bayes, Support Vector Machines	To Explore Preprocessing Techniques For Prediction Of Risk Of Readmission For Congestive Heart Failure Patients
Yang Xie et al	Regression Decision Tree	To Predict The Number Of Hospitalization Days Based On Health Insurance Claims Data Using Bagged Regression Trees
Peyman Rezaei Hachesu et al	Decision Tree, Support Vector Machines (SVM), And Artificial Neural Network (ANN)	To Predict The Hospitalization Days For Cardiac Patients Using Data Mining Techniques
April Morton et al	Multiple Linear Regression, Support Vector Machines, Multi-Task Learning, And Random Forests	To Compare Different Supervised Machine Learning Techniques For Predicting Short-Term In-Hospital Length Of Stay Among Diabetic Patients
Sajid Zaidi et al	Logistic Regression, Gradient Boosting With Logistic Regression, Randomforest, Svm And Ensemble	To Predi Ct Hospi Ta L Readmissions In The Medi Care Population
Tanuja et al	Decision Tree C4.5, Naïve Bayes Classifier, K-NN	To Compare Different Data Mining Techniques To Predict Hospital Length Of Stay
Parag C. Pendharkar et al	Support Vector Regression, Regression Tree	To Compare Different Machine Learning Techniques

6. Conclusion and Scope for Further Research

The paper has dealt with the background of a Predictive Analytics Tool, its domain and the methodology to calculate the number of days a patient is expected to be admitted to hospital using Charlson Co morbidity Index. With these basics, we will be focusing on the pre-implementation of the predictive analytic tool in our next paper. Future work includes an optimistic model that aims at employing the insurance data to predict the LOS of a patient.

7. Acknowledgement

I would like to take this opportunity to express my profound gratitude and deep regard to my guide, *Prof. M. Revathi, CSE, Kingston Engineering College*, for her exemplary guidance, valuable feedback and constant encouragement in completing this paper. Her valuable suggestions were of immense help in getting this work done. Working under her, was an extremely knowledgeable experience. Also, I would like to extend my sincere gratitude to my parents and my sister for their constant support and encouragement throughout my education.

References

[1] I. Duncan, “Mining health claims data for assessing patient risk,” in Data Mining: Foundations and Intelligent Paradigms, ser. Intelligent Systems Reference Library, D. E. Holmes and L. C. Jain, Eds. Springer Berlin Heidelberg, 2012, vol. 25, pp. 29–62.

[2] J. Donze, D. Aujesky, D. Williams, and J. L. Schnipper, “Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients: Derivation and Validation of a Prediction Model,” JAMA Internal Medicine, vol. 173, pp. 632–638, 2013.

[3] O. Hasan, D. O. Meltzer, S. A. Shaykevich, C. M. Bell, P. J. Kaboli, A. D. Auerbach, T. B. Wetterneck, V. M. Arora, J. Zhang, and J. L. Schnipper, “Hospital readmission in general medicine patients: A prediction model,” Journal of General Internal Medicine, vol. 25, pp. 211– 219, 2010.

[4] E. Coiera, Y. Wang, F. Magrabi, O. P. Concha, B. Gallego, and W. Runciman, “Predicting the cumulative risk of death during hospitalization by modeling weekend, weekday and diurnal mortality risks,” BMC Health Services Research, vol. 14, 2014.

[5] R. B. Cumming, D. Knutson, B. A. Cameron, and B. Derrick, “A comparative analysis of claims-based

Licensed Under Creative Commons Attribution CC BY

methods of health risk assessment for commercial populations," Final Report to the Society of Actuaries, 2002.
[6] [6] Y. Zhao, A. S. Ash, R. P. Ellis, J. Z. Ayanian, G. C. Pope, B. Bowen, and L. Weyuker, "Predicting pharmacy costs and other medical costs using diagnoses and drug claims," Medical Care, vol. 43, pp. 34–43, 2005.
[7] D. Bertsimas, M. V. Bjarnadottir, M. A. Kane, J. C. Kryder, R. Pandey, S. Vempala, and G. Wang, "Algorithmic prediction of health-care costs," Operations Research, vol. 56, pp. 1382–1392, 2008.
[8] K. Pietz, C. M. Ashton, M. McDonell, and N. P. Wray, "Predicting healthcare costs in a population of veterans affairs beneficiaries using diagnosis-based risk adjustment and self-reported health status," Medical Care, vol. 42, pp. 1027–1035, 2004.
[9] C. A. Powers, C. M. Meyer, M. C. Roebuck, and B. Vaziri, "Predictive modeling of total healthcare costs using pharmacy claims data - A comparison of alternative econometric cost modeling techniques," Medical Care, vol. 43, no. 11, pp. 1065–1072, 2005.
[10] (2014) How much do we spend on health? Australian Institute of Health and Welfare, Australian Government. [Online]. Available: http://www.aihw.gov.au/australias-health/2012/spending-on-health/
[11] H. G. Dove, I. Duncan, and A. Robb, "A prediction model for targeting low cost, high-risk members of managed care organizations," American Journal of Managed Care, vol. 9, no. 5, pp. 381–389, 2003.
[12] B. Fireman, J. Bartlett, and J. Selby, "Can disease management reduce health care costs by improving quality?" Health Affairs, vol. 23, no. 6, pp. 63–75, 2004.
[13] J. Polisena, D. Coyle, K. Coyle, and S. McGill, "Home telehealth for chronic disease management: A systematic review and an analysis of economic evaluations," International Journal of Technology Assessment in Health Care, vol. 25, pp. 339–349, 7 2009.
[14] "Report on government services 2013 volume 2: Health; community services; housing and homelessness," Steering Committee for the Review of Government Service Provision, Canberra: Productivity Commission, 2013.
[15] P. Brierley, D. Vogel, and R. Axelrod. (2014) Heritage provider network health prize round 1 milestone prize: How we did it - Team 'Market Makers'. [Online]. Available: http://www.heritagehealthprize.com/c/hhp/leaderboard/milestone1
[16] Wen-Yuan Jen "Mobile healthcare services in school-based health center", International Journal of Medical Informatics, Vol. 78, pp. 425–434, 2009.
[17] Ali Serhan Koyuncugil and Nermin Ozgulbas "Financial early warning system model and data mining application for risk detection", Expert Systems with Applications, Vol. 39, pp. 6238–6253, 2012.
[18] Sumana Sharma and Kweku Muata Osei-Bryson, "Framework for formal implementation of the business understanding phase of data mining projects", Expert Systems with Applications, Vol. 36, pp. 4114–4124, 2009.

Author Profile

S. Divya Meena is a Post-graduate student in the Computer Science Department, Kingston Engineering College, India. She received B.Tech degree in 2014 from VIT University, Vellore, India. Her research interests are Cloud Computing, Data Structures, Data Mining and Big Data.

M. Revathi, Assistant Professor, Department of Computer Science, Kingston Engineering College. She received her B.E degree in 2007 from SKP Engineering College. She then completed her M.Tech degree in 2014 at SRM University. Her area of interest includes Computer Networks Security, Wireless Networks, Sensor Networks, Data Structure and Cloud Computing. She has published three papers and has attended more than 15 conferences.

Licensed Under Creative Commons Attribution CC BY