ISSN: 0975-766X
CODEN: IJPTFI
Research Article

Available Online through
www.ijptonline.com

PROGNOSIS OF FUTURE HOSPITALIZATION - MINING HISTORIC HEALTH INSURANCE CLAIMS

Ganesh Jayakumar*[1], R. Rathna[2]
[1]PG Scholar, Department of Information Technology, Sathyabama University, Jeppiaar Nagar, Chennai-119, India.
[2]Assistant Professor, Department of Information Technology, Jeppiaar Nagar,
Sathyabama University, Chennai-119, India.
Emails: ganex27lin@gmail.com

Received on 20-04-2016 *Accepted on 15-05-2016*

Abstract

In the data mining world, the art of knowledge discovery in the large volume of data enables improvisation in various business needs. When the knowledge from mining the data termed into an action it proactively enriches the business satisfaction. In healthcare in order to augment the service and improve the business quality, predictive analysis plays a vibrant role. Under the various group of population health claims differ region to region and a common predictive model should be required to support numerous types of health claims. Decision Tree and Apriori algorithms are mainly used. It exactly finds out for how many number of days, the person is actually hospitalized. So it will be of great help to find out the number of days the person has spent in hospital. Through the predictive mechanism insurance companies can make their decision and provide the offers to improvise the business.

Keywords: Apriori Algorithm, Data Mining, Health Claims, Insurance, Prediction, etc.,

1. Introduction

Data mining is an approach for transferring huge volume of data into meaningful and useful knowledge. Those processed knowledge is transmitted into action which help to take business decisions. Healthcare administrators worldwide endeavoring to bring down the cost of care enhancing the nature of care given.

Hospitalization is the largest component of health expenditure. A method was developed, using large-scale health insurance Claims data, to predict the number of hospitalization days in a population. The proposed strategy performs well in the overall public and also in sub-populaces. Results indicate that this new model significantly improves the predictions over established baseline methods.

2. Related Work

The utilization of health risk assessment strategies based on available medical diagnosis codes from administrative claim information keeps on growing. The US government has implemented a process [1] that uses medical diagnosis codes to minimize or maximize the payments to Medicare Choice agents. Many states have used methods that use medical diagnosis codes to change the payments according to the health care plans for Medicaid entries. Diagnosis based methods of risk assessment have likewise been utilized by managers of health insurers in analyzing how employee commitments ought to fluctuate by decision of provider or health insurance plan. These insurers are increasingly using, diagnosis of risk assessment for provider profiling, case administration, payment, and rating [2]. Many Health Care organizations maintain population risk by applying disease management or case management programs. The targeting of specific events or diseases has more probability for risk because enrollees who were hospitalized [3] or have ≥1 specific risk diseases like asthma, cancer, heart diseases, diabetes, etc., experience higher claims costs than the average for the plan. However, there are many ways of trying to identify potentially high-cost patients including using diseases (such as AIDS or kidney failure) likely to require expensive treatments, previous hospital or emergency department utilization. Execution of 10th Revision of the International Statistical Classification of Disease and Related Health Problems, (ICD-10) coding system provides issues and challenges for using administrative data [4]. Utilizing this, a multistep operation has been set up to develop ICD-10 coding startegies to define Charlson and Elixhauser comorbidities in the available administrative data and assess the performance of the resulting strategies. The ICD-9-CM adaptation [5] of the conventional Charlson comorbidity score has been a useful resource for health service providers and researchers. With the newly introduced ICD-10 coding worldwide, an ICD-10 version of the Deyo adaptation was introduced as well as validated using people statistics based hospital data from Australia. The new strategy of ICD-10-AM was a modification of ICD-9-CM (Australian modification). After applying mapping calculation the results were utilized to add to an underlying interpretation, these codes were physically analyzed by the coding specialists and a general doctor for face legitimacy. Because of the region specific nature of ICD-10 system, the goal of the proposed methodology is to keep much of the translated code at a simple three digit level for increasing general usage of the new index. To build the set-membership, a new strategy is applied in this proposed work. Initially it is done based on bilinear group assumptions. While applying to the case where Φ is a range of integers, the new strategy requires to swap $O(\ k\log k - \log\log k\ )$ group elements. Instead by using the commitment schemes where the dependency is on RSA, there are solutions to this problem which needs only a fixed number of RSA-group elements to be swapped between the prover and verifier.

3. Existing System

The process of reimbursing health oriented services in return for a fixed monthly premium for those who enrolled for Health Insurance policies compels the Helath Insurers to critically evaluate them before reimbursing. Poor danger measure could bring about surpassing the budget economically which leaves the Insurers in a financially crucial state. Most of the Insurers are using Log-linear analysis algorithm for accumulating the numeric retreived from the admission process and it maps the actual calculated claim levels with that of the customer level like the earlier cost information [6]. But this technique restricts heavy volume of data such as Claim Data, which could reduce the percentage of prediction.

4. Proposed Model

A Proposal has been made to develop a generalized model (Fig.1) that would guess the total number of days admitted in a hospital during a year for people (who have enrolled for health insurance) from a overall public [7], using extensive scale medical coverage claims information. Since protection claims have solid financial attributes [8], their power in anticipating clinical targets, for example, hospitalizations, are from time to time researched. A hybrid combination of Decision tree algorithm and Apriori algorithm is proposed to be used in this work. The most often used Decision trees are powerful tools for classification and prediction. Decision trees represent rules [9], which can be understood by humans and used in knowledge system such as making persuading fake confirmation of produced information in cipher texts such that outside coercers are fulfilled. It supports large volume of data. The Insurance claim department [10] gains the knowledge of this and helps in deciding for which patient, which type of Insurance is better to provide for. Because which type of diseases are faced [11] often in hospital records can be easily mined using this technique.

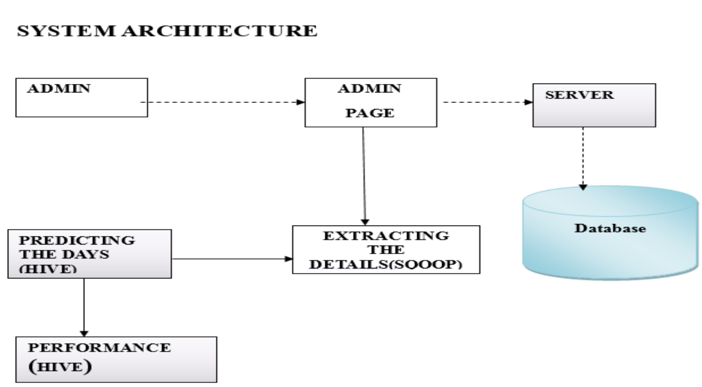

Fig 1. Architecture Diagram

4.1. Patient registering details in Hospital

In this module we design the UI for the project. These windows are used to send a message from one peer to another. Those who use the application need to login through the login page using the created GUI. It acts as an interface to connect User and the actual Database. This login screen is used by the user to input user name and password. Then this module does the validation for verifying whether the user is a registered user. If registered then it gives access (Fig.2).

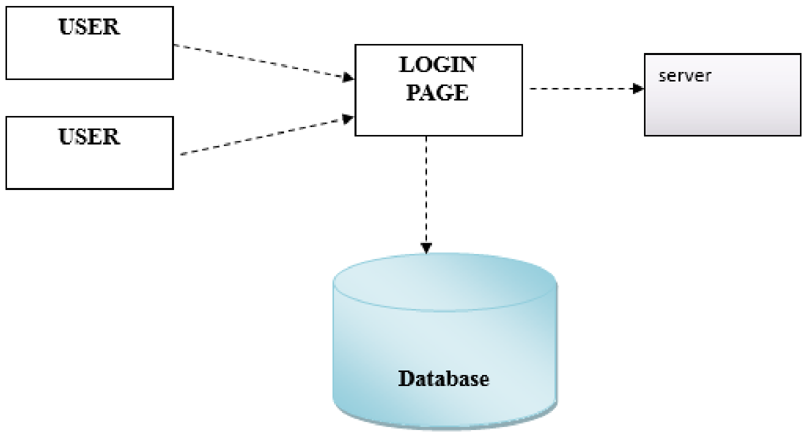

Fig 2. Login Page.

4.2. Record Maintenance

When the user requests for the doctor's appointment, the hospital admin allocate a slot for appointment with the concern doctor (Fig.3).

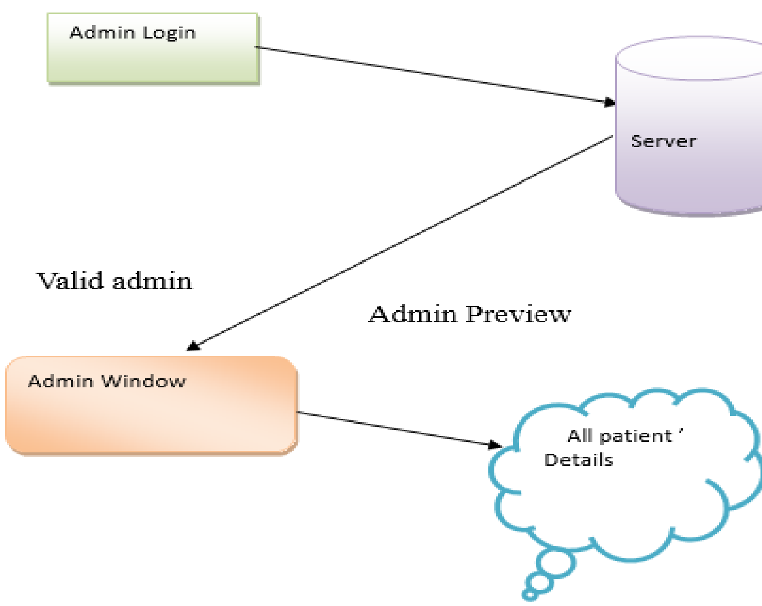

Fig 3. Admin Page.

In this module the admin can fix the appointment according to the request and block doctor's calendar for further appointment registration.

4.3. Extracting the records through Sqoop.

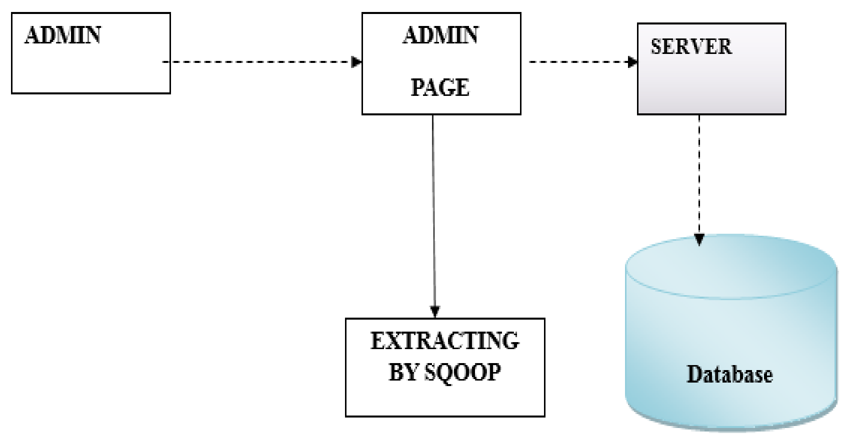

Fig 4. Record Extraction using SQOOP.

Once the patient have the appointment with the requested doctor, the patient will have the consultation accordingly. Then the payment will be made through either their insurance or cash (Fig.4). At the same time the data about this diagnosis gets stored using MySql [12]. From the MySql database, the data is dumped into Hive database through the Sqoop tool.

4.4. Predicting the number of days

Once the data about the diagnosis of the patient and the disease data are available in MySql database, Hive environment gets ready with various data nodes to handle the historic information [13] with the past few years of data for comparison. The Job tracker then interacts with Name node and schedules the tasks according to the decision made and resource availability.

4.5. Performance Evaluation

The main performance indicator is referred to as the root-mean-square error (RMSE), and is the root-mean-square of the difference between the logarithm of the estimated DIH and the logarithm of the true number of days. This performance measure is done through hive map reduce mechanism, by segregating various partitioning of the big volume of data.

5. Conclusion and Future Enhancement

The Experiment is done by gathering various historical health claims and patient history and predict future hospitalization [Fig.5].

Fig.5. The Prediction.

The conventional technique using Java MapReduce program for structured, semi-structured, and unstructured data has been done. The scripting approach for MapReduce to process structured and semi structured data using Pig is executed. The Hive Query Language (HiveQL) is used for MapReduce to process structured data using Hive. Through the MapReduce and parallel procession mechanism huge volume of data is processed and predicting samples are derived.

Days in hospital can be further categorized into various graded categories, as this is a good predictor for targeted interventions from an insurance perspective. A combination of the optimal set of candidate models usually predicts more accurately using the proposed methodology.

Acknowledgement

We the authors are very much grateful to the management of Sathyabama University for providing us with all the necessary resources and space for carrying out this work.

References

1. 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 BerlinHeidelberg, 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. Hude Quan et.al., "Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries", American Journal of Epidemiology, pp. 1-7, 2011.
6. R. B. Cumming, D. Knutson, B. A. Cameron, and B. Derrick, "A comparative analysis of claims-based methods of health risk assessment for commercial populations," Final Report to the Society of Actuaries, 2002.
7. 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.
8. 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.
9. 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.
10. 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.
11. Mary Gladence, L.,Ravi, T.,Karthi, M, "An enhanced method for detecting congestive heart failure-Automatic Classifier", Proceedings of 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICACCCT 2014.
12. 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.

13. S.Ancy, R.Kumar, R.Asokan, R.Subhashini, "Prediction of onset of south west monsoon using multiple regression", Proceedings of ICCCS 2014 - IEEE International Conference on Computer Communication and Systems. Pp.170-175. Feb 2014.

Corresponding Author:

Ganesh Jayakumar*,

Emails: *ganex27lin@gmail.com*