

Providing Better Medical Healthcare Services using Datamining Techniques

Geetha Guttikonda
Assistant Professor
Kanuru, VRSEC

Madhavi Katamaneni
Assistant Professor
Kanuru, VRSEC

Prasanna Komara
M.Tech 2nd Year
Kanuru, VRSEC

Abstract - To provide better medical healthcare services to the people, the healthcare centers are need to predict the patient's revisit. And we can also find the diseases of patients who can revisit the healthcare center. To predict the patient's revisit we can apply classification models like Decision tree, Naïve Bayesian. We can find the accuracy of patient's revisit by using these models. To find the diseases of revisited patients we can apply sequential pattern analysis. This type of predictions used in healthcare centers to create awareness for people to revisit.

I. INTRODUCTION

Data Mining is a standout amongst the most indispensable and rousing range of research with the goal of discovering significant data from tremendous information sets. In present period, Data Mining is getting to be well known in medicinal services field on the grounds that there is a need of productive diagnostic approach for distinguishing obscure and important data in wellbeing information. In wellbeing industry, Data Mining gives a few benefits, for example, discovery of the misrepresentation in medical coverage, accessibility of medicinal answer for the patients at lower cost, recognition of reasons for sicknesses and distinguishing proof of therapeutic treatment strategies. It too offers the medicinal services analysts for making productive human services strategies, some assistance with constructing drug proposal frameworks, creating wellbeing profiles of people and so on [1].

Progressing and keeping up a not too bad broad prosperity is a creating concern both of national and of neighborhood governing bodies of India as in various countries, and it contains practices which add to making general prosperity approach and passing on human administrations organizations. General wellbeing area ought to assume control also, deal with the medicinal services administrations which have been disregarded by private wellbeing part, and this perspective ought to be reflected in general wellbeing strategy. In India general wellbeing part still stays experiencing personality emergency.

Public healthcare centers focuses need to attempt to determine these issues. At the full scale level, for development in administrative proficiency, it is essential for public healthcare centers to set up and execute adaptable arrangement that reflects wellbeing request by neighborhood inhabitants and designates spending plan and labor as indicated by the wellbeing request. At the miniaturized scale level, the neighborhood governments need to do their best for wellbeing change of the neighborhood group by arranging and executing medicinal services administration programs and organizations, urging neighborhood occupants to visit general wellbeing focuses and ensuring adequate social insurance assets thus on[2]. Since the area restorative administrations associations are driven by open social insurance focuses, it is key to have open human services focuses expect their parts and perform their abilities well [2].

Public healthcare centers focuses plan to offer generous social insurance administrations to neighborhood inhabitants by building up human services approach suitable for nearby conditions. In this way, gets ready for this social insurance approach (e.g., affiliation, work, workplaces, sorts of rigging, monetary arrangement, and progression courses of action) can be made through the estimation of prosperity interest, which can be assessed by anticipating patient's arrival to as a starting stage. Along these lines, specially, thinking about the likelihood of patient' return to is crucial to setting up suitable social insurance arrangement, fitting to neighborhood conditions.

Furthermore, public healthcare centers focuses attempt to give data about systems to ailment counteractive action what's more, treatment with neighborhood residents. To give better restorative organizations with respect to balancing activity and treatment of most unremitting ailments, it is required to light up patients of wellbeing focused measures as showed by their anticipated diseases. This study looked to discover responses to the accompanying inquiries: 1) Can we anticipate whether a patient will return to a public healthcare center? 2) Can we foresee illnesses of patients who return to the inside? Answers to these inquiries will be useful in making strides administrative effectiveness of the inside and giving a superior restorative support of the patients by recommending safety oriented measure to them.

As a way to give a response to the patient's revisit, we applied three classification algorithms (i.e., decision tree [DT] and Naïve Bayesian [NB]) method proposed in this study to building classification models. To predict the diseases of revisit patients, we performed sequential pattern analysis with a reason to recognize predictable ailments of returning to patients.

II. PUBLIC HEALTHCARE SYSTEM IN INDIA

PHC is a referral unit for six sub-focuses. All PHCs give outpatient administrations; a larger part has four to six in - quiet beds. By standards they have one restorative officer, 14 Para - therapeutic and other supporting staff. At the national level there are more than a sufficient number of PHCs and specialists posted at PHCs however the appropriation crosswise over states is uneven; there are no utilitarian PHCs in numerous remote regions in desperate need of social insurance.

Another National Health Policy was reported in 2002, which recognized that general society medicinal services framework is horribly lacking on different fronts and asset designations are by and large deficient. While this arrangement expressed objectives like "expansion usage of general wellbeing offices from current level of under 20% to more than 75%", no relating vast scale measures for reviving and fortifying the incapacitated general wellbeing framework were arranged.

The aggregate estimation of the wellbeing segment in India today is every year over Rs.150, 000 crore or US\$ 34 billion. This works out to about Rs .1500 per capita which is 6 for each penny of GDP (Table 1.1). Be that as it may, of this just 15 for each penny is openly financed, 4 for each penny is from social protection, 1 for each penny private protection and the remaining 80 for every penny is spent out of individual resources.(85 per penny of which goes to the private segment).

There is a requirement for a subjective increment in assets for open wellbeing to the level of 3% of GDP in the short term, moving to 5% of the GDP in medium term. These raised open accounts for the general wellbeing framework could be raised through general tax collection alongside different types of exceptional tax assessment what's more, taxes for wellbeing security. Also, finishing sponsorship of the private therapeutic part and successfully burdened this segment, particularly its upper end; an extraordinary wellbeing security tax on all money related exchanges including worldwide exchanges above a specific level; and particular tax assessment of commercial ventures with negative wellbeing effects are some different measures that could be received.

III. LITERATURE SURVEY

Extending use of data mining strategies can be found in a wide variety of areas, for instance, cash, retail showcases, tele correspondence, therapeutic domain, in this way on [3-7]. Past research has demonstrated that information mining systems can be utilized to evoke undiscovered valuable learning from huge therapeutic datasets [8,9]. This area audits past examination which uses characterization or successive example investigation for different undertakings in medicinal or social insurance space.

Choi et al. [10] built up a crossover model for foreseeing 5-year survival rates for bosom disease by consolidating the manufactured neural system and Bayesian system. Phillips-Wren et al. [11] anticipated whether a lung disease patient will visit the restorative oncologist or not by utilizing choice tree, logistic relapse, and simulated neural system. Polat et al. [12] grouped sound and strong infected subjects by utilizing choice tree. Ture et al. [13] anticipated the illness free survival in bosom malignancy patients by utilizing 6 choice tree calculations – grouping and relapse tree, chi-squared programmed communication indicator, brisk, fair-minded and proficient measurable tree, iterative dichotomiser 3(ID3), business 4.5 (C4.5), business 5.0 and cox relapse. Kang et al. [14] made 2 counterfeit neural system models and 2 characterizations and relapse tree to anticipate both the total of specialist's office charges and the measure of costs paid by the protection of malignancy patients.

Patterns in health care domain include the common patterns in paths followed by patients in hospitals, patterns observed in symptoms of a particular disease, patterns in daily activity and health data. Exarchos et al. [15] analyzed and classified the proteins by using protein sequence. That is, they isolated progressive case of proteins, which were then used to aggregate the dark proteins. Chiang et al. [16] extricated collaboration

patterns between qualities acquired from biomedical records. To be particular, this study added to another consecutive example mining technique to mine significant standards that portray the sorts of morphological components that can show up previously, then after the fact the name of quality in records. Asker [17] distinguished patients' ailment on the premise of their successive manifestations. Right when patients have one of the anticipated sicknesses, each of these afflictions may be insisted by specific progressive indications. Lin et al. [18] added to a progressive data mining strategy which is useful to orchestrate tolerant thought activities, to diminish sharpen assortments, and to minimize delays in prescriptions with the end goal of empowering the endless change of designating more appropriate clinical approaches to brain stroke patients.

IV. RESEARCH ARCHITECTURE

Patient's data is collected from Andhra Hospitals, Vijayawada in 2015 and then following steps are applied according to architecture design as shown in Fig.1.

1) Pre-processing

We apply some pre-processing techniques on that like remove unnecessary data, filling up the empty fields and so on. We create two pre-processing models for the change from ostensible qualities into binominal characteristics (every trait just has two unique values now) and a second one for changing every single binominal property into numerical ones (comprising of 0 and 1).

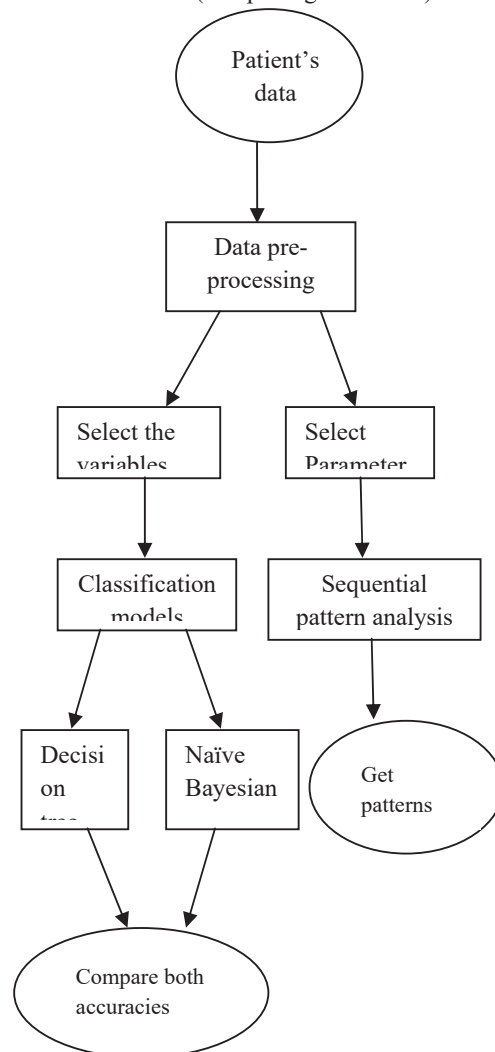


Fig.1: architecture design for comparison

This is extremely regular pre-processing chain for the change of information sets containing ostensible properties into information sets comprising of numerical qualities as it were.

2) Select the variables from patient's data

Numerous variables have been utilized as a part of the past examines to foresee whether patients will return to an open medicinal services focus or not. Phillips-Wren et al. [11] used both sociodemographic moreover, clinical trademark data in their study to envision whether a lung danger patient will visit the restorative oncologist or not. These variables demonstrate quiet conditions, demographics, and medicines. They have been accepted in different human services related studies [11, 19, and 20].

For classification, we can choose 10 variables from our data (i.e., Patient ID, gender, age, height, weight, distance, systolic pressure, diastolic pressure, name of disease, prescription time).

3) Apply classification models

To perform our patient's revisit task, we used decision tree, naïve Bayesian models which is widely used for various medical data analysis. We assessed 10 information variables utilizing Gain Ratio characteristic evaluator taking into account ranker pursuit technique, to choose more powerful variables while foreseeing the objective variable.

Decision Tree:

To perform decision tree we are using ID3 algorithm. The fundamental thought of ID3 calculation is to make a decision tree of given set, by utilizing top- down insatiable pursuit to check every characteristic at each tree hub. After generating decision tree, we can find the confusion matrix from values of generated decision tree. From that confusion matrix we calculate the accuracy of patient's revisit.

Naïve Bayesian:

Naive Bayes learners and classifiers can be amazingly quick contrasted with additional refined strategies. The decoupling of the class restrictive component dispersions implies that every conveyance can be autonomously evaluated as a one dimensional circulation. This thusly eases issues coming from the scourge of dimensionality. The predictions can be contrasted with the class values in the test dataset and a characterization exactness can be computed as a precision proportion somewhere around 0% and 100%.

4) Sequential pattern analysis

For predicting diseases of revisited patients we are using generalized sequential pattern (GSP) algorithm can be used. The Generalized Sequence Pattern calculation was made from a less difficult calculation for mining groupings, yet it has some additional extravagant accessories included so it can be more adaptable for various circumstances. To apply this algorithm we can convert the patient's data into numeric. After converting numeric we can easily get patterns.

V. RESULTS

We can apply model for classification techniques and then performance in java application. Finally we get accuracy from confusion matrix which can be tabulated below. By using decision tree we get 83.28% accuracy and from naïve Bayesian we get 79.25% accuracy. We compare two algorithms accuracies to know which model can take better decision in patient's revisit. Decision tree is the best algorithm in patient's revisit than compare to naïve Bayesian.

Table 1: Results after applying classification Models:

Algorithm	Accuracy
Decision tree	83.28
Naive Bayesian	79.25

Sequential patterns from patient's data can be displayed in below table. We can take minimum support count 2% and minimum confidence 12%. Since it is difficult to determine consecutive examples with high support from true information in restorative area and the best possible least backing depends both on the attributes of

issues to be tackled and on the strategy of the establishment which will utilize the successive examples, we set the base backing to a low esteem, like the one which different analysts have set.

Table 2: Results from sequential pattern analysis

SequenceID	Patterns
1	USG abdomen,LFT,TC&DC ->cholelithiasis
2	USG abdomen,Endoscopy,2D echo, LFT, TC&DC -> cholelithiasis
3	USG abdomen,2D echo, MRCP, ERCP -> cholelithiasis
4	2Decho,LFT,malariaantigen,Dengue,TC&DC ->Viral fever With thrombocytopenia
5	TC&DC,LFT,2D echo, Dengue -> Viral fever With thrombocytopenia
6	CT report, Pathology report -> Carcinoma right gingival buckle sucks

VI. CONCLUSION

Grouping models and successive examples can help open medicinal services focuses plan and actualize social insurance administration projects and organizations that are more proper to neighborhood occupants, urging them to return to general wellbeing focuses. From the results we can get decision tree is best for patient's revisit. But naïve Bayesian also best for another medical decision. By observing these results we can create awareness to the people to go to the public healthcare center frequently for regular check up. And also we can tell the people which are having dangerous diseases till they don't know. So they are alert and take precautions before attack.

REFERENCES

- [1] H.C.Koh and G.Tan, "Data Mining Application in Healthcare", Journal of Healthcare Information Management, vol. 19, no. 2, (2005).
- [2] Nam CH. Policy development on health administration system in the era of local autonomous government. J Korean Soc Health Educ Promot 1996; 16: 101-126.
- [3] Chae YM, Kim HS, Tark KC, Park HJ, Ho SH. Analysis of healthcare quality indicator using data mining and decision support system. Expert Syst Appl 2003; 24: 167-172.
- [4] Han HK, Kim HS, Sohn SY. Sequential association rules for forecasting failure patterns of aircrafts in Korean airforce. Expert Syst Appl 2009; 36: 1129-1133.
- [5] Hung SY, Yen DC, Wang HY. Applying data mining to telecom churn management. Expert Syst Appl 2006; 31: 515-524.
- [6] Wu LG, Lee JX, Huang HD, Liu BJ, Horng JT. An expert system to predict protein thermostability using decision tree. Expert Syst Appl 2009; 36: 9007-9014.
- [7] Yeh IC, Lien CH. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Syst Appl 2009; 36: 2473-2480.
- [8] Lee SM, Park RW. Basic concepts and principles of data mining in clinical practice. J Korean Soc Med Inform 2009; 15: 175-189.
- [9] Silva A, Cortez P, Santos MF, Gomes L, Neves J. Mortality assessment in intensive care units via adverse events using artificial neural networks. Artif Intell Med 2006; 36: 223-234.
- [10] Choi JP, Han TH, Park RW. A hybrid Bayesian network model for predicting breast cancer prognosis. J Korean Soc Med Inform 2009; 15: 49-57.
- [11] Phillips-Wren G, Sharkey P, Dy SM. Mining lung cancer patient data to assess healthcare resource utilization. Expert Syst Appl 2008; 35: 1611-1619.
- [12] Polat K, Kara SI, Guven A, Gunes S. Usage of class dependency based feature selection and fuzzy weighted pre-processing methods on classification of macular disease. Expert Syst Appl 2009; 36: 2584-2591.
- [13] Ture M, Tokatli F, Omurlu IK. The comparisons of prognostic indexes using data mining techniques and Cox regression analysis in the breast cancer data. Expert Syst Appl 2009; 36: 8247-8254.
- [14] Kang JO, Chung SH, Suh YM. Prediction of hospital charges for the cancer patients with data mining techniques. J Korean Soc Med Inform 2009; 15: 13-23.
- [15] Exarchos TP, Papaloukas C, Lampros C, Fotiadis DI. Mining sequential patterns for protein fold recognition. J Biomed Inform 2008; 41: 165-179.
- [16] Chiang JH, Liu HS, Chao SY, Chen CY. Discovering gene-gene relations from sequential sentence patterns in biomedical literature. Expert Syst Appl 2007; 33: 1036-1041.
- [17] Lasker GE. Application of sequential pattern-recognition technique to medical diagnostics. Int J Biomed Comput 1970; 1: 173-186.
- [18] Lin F, Chou S, Pan S, Chen Y. Mining time dependency patterns in clinical pathways. Int J Med Inform 2001; 62:11-25.
- [19] Cooper GS, Virnig B, Klabunde CN, Schussler N, Freeman J, Warren JL. Use of SEER-Medicare data for measuring cancer surgery. Med Care 2002; 40(8 Suppl): IV- 43- IV-48.

- [20] Earle CC, Nattinger AB, Potosky AL, Lang K, Mallick R, Berger M, Warren JL. Identifying cancer relapse using SEER-Medicare data. *Med Care* 2002; 40(8 Suppl): IV- 75-IV 81.
- [21] Oh HS. Health behavior and utilization of the public health center in rural adults. *J Honam Univ Acad* 2008; 29: 361-376.
- [22] Yoo HR. Implementing a smoking cessation clinic at a public health center in Korea: evaluating the outcomes and the smokers' perceptions. *J Korean Acad Public Health Nurs* 2008; 22: 62-73.
- [23] Yoon HS, Lee HY, Lee SK. Factors associated with the use of health promotion program: Seoul community health center. *Health Soc Welf Rev* 2008; 28: 157-184.