International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 6; Issue: 12; December -2017

Comparative Analysis of Accuracy and Precision in HealthCare Customer Relationship Management System

Naveen Kumar B
Department of Computer Science & Engineering
Sambhram Institute of Technology
Bangalore- Karnataka
navkan24@gmail.com

Mohd.Shah Alam
Department of Computer Science & Engineering
CSJMU
Kanpur-Utter Pradesh.
shahalam13@gmail.com.

***Abstract*—India has a vast health care system, but there remain many differences in quality between rural and urban areas as well as between public and private health care. To provide effective in contributing to better health throughout the entire population and safeguarding patient's dignity, confidentiality, atomicity and being sensitive to the specific needs and vulnerabilities of all population group, I prosed a prediction healthcare system.This paper focuses on comparing the accuracy and precision of different prediction system with proposal prediction system which highlights on a prediction of diseases based on currently available patients data. This localization proposal would bring a system in where a healthcare prediction system and patients are the initial point of contact and helps to provide a different facet to Indian medical scenario using technological advancement.**

***IndexTerms*—Healthcare, CRM, Healthcare prediction system Portal.**

I. Introduction

India is the second largest population in the world. Almost two-third of the people hassuffering from one or other diseases. To monitor health of every citizen, there is a need of healthcare customer relationship management (CRM).

The healthcare CRM will act as a mediator between healthcare organization and patients who is suffering from various diseases. The healthcare CRM was broadly classified into two parts, one is for healthcare organization to stay in contact with the people who are suffering from many diseases and the other one is for healthcare organization to stay in contact with referring organizations.

The healthcare systems automatically collect all information and harmonize data from Laboratory Information Systems (LIS), billing, payer, supply and other sources.Hospitals and specialty clinics use a healthcare CRMto maintain relationships with primary care physicians or other referring healthcare organizations.This personal information is used by healthcare professionals in order to predict patient's future issues and deliver messages however and wherever they are most relevant to that individual.

To do so successfully, healthcare marketers need technology that facilitates the creation of personalized messaging and automated engagement efforts. Marketing technology can help teams create integrated, Omni channel campaigns by incorporating analytics, strategy, and creativity. It also provides insights into customer data, giving healthcare marketers the tools to keep up with their evolving customer base. With these solutions, healthcare marketers can deliver consistent, timely, and relevant messages that acquire, retain, and engage patients.

II. Literature Survey

The importance of CRM in health care system has been felt of paramount importance and many works have been done in the literature. A review of some notable work is presented in this section.

[1] M C S Geetha, Elizabeth Shanti et.al was focused to apply and analyze the commonly used classification algorithms on medical data set that helps to predict heart diseases in future. In their research, the goals of prediction can attain by using following mining techniques like:
(1) Data Classification
(2) Regression
(3) Clustering
(4) Summarization
(5) Dependency Modeling
(6) Change and Deviation detection and they used WEKA tool to evaluate the performance using confusion matrix.
The confusion Matrix reveals the parameters such as accuracy, sensitivity and specific measures and they have used evaluation criteria-based timing to build model, correctly classified instances, incorrectly classified instances and predictive accuracy. Some of the classification algorithms

such as J48, CART and REPTREE was used to predict the heart diseases and these algorithms provide the better accuracy compared to other algorithms. In their research, they conducted experiments on 11 attributes to predict heart diseases. Through the experimental results they concluded that J48, SIMPLE CART, REPTREE algorithm prove the best classification techniques in terms of accuracy. Finally, they concluded that predictive accuracy determined by algorithms JS-48, REPTREE and Bayes Net are consistent indicator to predict heart diseases.

[2] Gunarathne W H S.D, Perara K D M et.al proposed an automated machine learning solution to detect chronic kidney disease by using 14 attributes out of 25 attributes present inside the dataset. In their research, the dataset is taken from machine learning UCI repository. This dataset contains 400 patient records. In their research, they buildclassification models with different classification algorithms. By using these classification models, they predict the chronic kidney diseases status of the patients. The accuracy results are compared against the models and choose the best model which gives better accuracy. The best model was used to predict the chronic kidney diseases status of the patients. Early detection of the diseases will give the benefits to the patients and doctors in providing early treatments and reducing the distribution of diseases progression. In their research, they have used different algorithms called (1) multiclass decision forest gives accuracy of 99.1%. (2) Multiclass decision jungle gives accuracy of 96.6%. (3) Multiclass logistic regression gives accuracy of 95.0%. (4) Multiclass neural network gives accuracy of 97.5%. In their research, they used cross industry standard process for data mining as research methodology. In their research, multiclass decision forest algorithm performed best with an accuracy of 99%.The advantages of their research is that prediction process is less time consuming.

[3] Ms. Tejaswini U Mane was proposed a system that will predict the heart diseases based on some parameters like chest pain, cholesterol, age, resting BP and many more. In their research, they improved k-means clustering algorithm to some extent to increase the prediction of heart diseases and also they used classification algorithm called Decision tree algorithm that is ID3, which will be used for hybrid approach. They concluded that improved k-means algorithm showing accuracy better than single k-means based on central selection. They have proved by conducting experiments that data size of records 20000, improved k-means accuracy will be 96.73% but k-means gives 92.7% accuracy and also improved k-means will give better time complexity and good clustering.

[4] P Amrutha Valli, K R S Pravallika et. Al proposed a model for predicting the future trend of diseases using twitter data. In their research, they build single linear regression model by using tweets data. Live streaming was done for finding the tweets from forest gump website. The accuracy was less when they use incomplete data. To improve the accuracy ridge regularization method is used. The ridge regularization method is used to eliminate the prediction error.

In their research, data is collected from Podargos Company, which is a data service company. They have collected the data for five years. By comparing stored information and the actual data, they derived one graph and applied ridge regression method to eliminate prediction error.

Finally, they build one model to predict death rates for the selected states for the future years. The model gives accuracy of 80%.

[5] A S Yuksel,S F Cankaya et.al was proposed a system that will not only minimized the circulation of spam in email messaging systems, but will also put forward a step in the direction of trust identities and accountability in email infrastructure to overcome the spam problem. In their research, theprototype of the system has been developed on the Microsoft azure platform and thebehavior of emailaddress has been analyzed. By using support vector machine algorithm and Decision tree algorithm.

In order to correctly modeling the behavior of servers they used Pearson correlation test. This will find the correlation between the frequencies of email messages. They have considered two important parameters.

(1) False positive rate (2) accuracy, to measure the performance of spam classification. Finally, their research of study analysis reveals that methods to prevent and filter spam through DNSBL, white lists and antispam filtering are not enough, So that they introduced trust identities that will transform the current email systems into a more secure email infrastructure.

[6] P SureshKumar and S Pranavi was focused on diabetes dataset to predict types of diabetes that is type1, type2 or type3 and to check suitable machine learning algorithm such as random forest, support vector machine, LDA, CART, K-NN to give accurate predictions. They consider some of the evaluation metrics like: (1) Kappa (2) Precision (3) recall (4) sensitivity (5) specificity. The data set are collected from diagnosis lab in Warangal TS, India. They consider only 15 attributes inside that dataset.

All these metrics can be defined with the support of four cardinalities of confusion matrix. Such as (1) true-negative (TN) (2) true positive (TP) (3) false negative (FN) (4) false positive (FP).

The precision can obtain by using formula

$$Precision=TP/ (TP+FP)$$

The accuracy can obtain by using formula

$$Accuracy= (TP+TN)/(TP+TN+FP+FN)$$

The Kappa can obtain by using formula

Kappa= (accuracy-expaccuracy)/(1-expaccuracy)

The sensitivity can obtain by using formula

Sensitivity=TP/ (TP+FN)

The specificity can obtain by using formula

Specificity=TN/ (FP+TN)

Finally, they concluded that random forest algorithm gives better accuracy compared to all other algorithms like LDA, SVM, CART and K-NN. The accuracy given by random forest is 99%.The precision value is 1 and recall value is 1.

[7]. MIN CHEN, YIXUE HAO et.al proposes a new convolutional neural network based on multimodal diseases risk prediction (CNN-MDRP) algorithm. This algorithm is applicable for both structured and unstructured data. To overcome the difficulty of incomplete data, they have used latent factor model to reconstruct the missing data. They have taken one case study on regional chronic diseases of cerebral infarction. In their research, they have used performance metrics in order to evaluate performance. The performance metrics are (1) True Positive (TP). (2)False Positive (FP) (3) TrueNegative (TN) (4) FalseNegative (FN).

By using these metrics, they calculated four measurements such as

(1) Accuracy

Accuracy= (TP+TN)/(TP+FP+TN+FN)

(2) Precision

Precision=TP/ (TP+FP)

(3) Recall

Recall=TP/ (TP+FP)

(4) FI-measure

Fi-measure= (2*precision*recall)/(precision recall)

In their research, the goal is to predict whether a patient is amongst the cerebral infraction high-risk or low risk of cerebral. They have taken 760 patients records as dataset. In that they are concentrated on 79 features. They have used Naive Bayesian, K-nearest neighbor and decision tree algorithm (along with CNN-MDRP) to predict the risk of cerebral infarction diseases. The prediction accuracy of proposed algorithm reaches 94.8% with a convergence speed greater than any other algorithms.

[8] Rashmi G Saboji, Prem Kumar Ramesh proposed a solution for heart disease prediction based on attributes. Here, Heart diseases are a general name for a variety of diseases, whose symptoms may vary depending on the specific type of heart diseases. In their research, first they have collected the health care data of diseases and they stored the collected data in HADOOP distributed file system and then perform some processing on spark.

After storing data in HDFS, they applied random forest algorithm. Finally, by observing the result of random forest algorithm, they analyze the performance in terms of accuracy and error rate. The dataset used in their research is collected from UCI machine learning which contains different important parameters like ECR, Cholesterol, Chest Pain, Fasting sugar etc. The Heart patients are categorized into two classes that is, Class 0 and Class 1. The Class 0 indicates absence of heart disease and Class 1 indicates presence of heart disease.

In their research, they applied random forest algorithm in spark frame work for predicting heart diseases and they are able to achieve 98%accuracy and they concluded that random forest algorithm gives better accuracy compared to Naïve Bayes prediction accuracy.

[9] Sonu S Ravi Ranjan et. al was proposed a system that predict if a person has Parkinson's disease or not. The Parkinson's disease is a central nervous system disorder which is characterized by rigidity of muscles, slow slurred speech and vocal cord disorder in the early stage. The main aim of thesystem is to apply suitable data mining and analytic algorithms to voice recording datasets and obtained results. By using these results, they can predict Parkinson's disease. In their research, first collect the voice of patients through recording and then it is converted into various attributes of voice like jitter, shimmer, frequency by using PRATT script and then they are stored in dataset. Classification Algorithm called logistic regression in order to classify the data on the dataset. Decision tree algorithm was used to predict Parkinson's disease. Decision tree gives around 100% accuracy without feature selection. In their research, the dataset was taken from UCI machine learning repository. In their research-means clustering algorithm was used to identify the low, moderate and high condition of the diseases.

If we use feature selection, Decision tree gives about 88% to 94% accuracy. They have used extra tree classifier for knowing the feature scores.

[10] Alexandra L Heureux, Katarina Grolinger et.al was proposed in his research that volume, velocity, veracity poses the challenges to the Big Data and gives the solutions for each of these challenges. The volume includes challenges like processing performance, curse of modularity, class imbalance, curse of dimensionality, feature engineering, Non-Linearity, Bonferroni's principle and variance and Bias. The velocity includes the challenges like Data availability, Real-time processing/streaming, and concept drift. The Variety includes the challenges like Data Locality, Data Heterogeneity, Dirty and noisy data. The veracity includes challenges like Data provenance, Data Uncertainty. If these challenges are not handled properly. Then it will throw threat to the big data.

For all these challenges, In their research gave solutions to handle each volume, velocity, variety or veracity challenges further in their research is focused on applying other unstructured multioutput regression models and including them into the GCRF model and shows the assessment of stroke risk for a given patients data in real time.

[11] Milos Jordan ski, Milos Radovic et.al, proposed a machine learning based approach for calculation of wall shear stress distribution, which may play an important role in mechanisms related to initiation and development of atherosclerosis. In their research, they have used multivariate linear regression, multilayer perception neural network and Gaussian conditional random fields in order to capture relationships between geometric parameters, density, viscosity and velocity.

In their research, they are going to test two hypotheses.
(1) At different cycle time points they plotted WSS distributions for geometrically parameterized models of AAA and Carotid bifurcation can be modeled by using machine learning approaches.
(2) For capturing spatial relationships, they use Gaussian Condition Random Fields (GCRF) model that may lead to improvement in overall WSS distributions prediction accuracy.

The results obtained from simulations showed that GCRF was able to improve accuracy on both AAA and Carotid Bifurcation Models.

[12] Shobana V, N Kumar proposed a recommender system in healthcare. It also particularly concentrates on prediction and diagnosis of thyroid disorders in women by making appropriate classification and clustering they predict the disorders in thyroid hormone and recommending appropriate treatments. In their research, they have used MAHOUT algorithm for the prediction of thyroid in women's. The Apache MAHOUT is a library of machine learning algorithms for HADOOP.

[13] Peter k Ghavami, Kailash C et.al, proposes a prognostics engine using multiple models to predict patient physiological status. In their research, the multimodal ensemble method is a combination of four methods that is they have used four different methods to predict patient's health status. The four models are (1) Probabilistic Neural Network (2) Generalized feed forward multi-layer MLP with Levenberg-marquardt(LM) (3)MLP trained with LM (4)Support Vector Machine .In their research, They have concluded that by combining multiple models they can improve the accuracy.

[14] Thanga Prasad S, Sangavi et .al was proposed a system to predict and classify the type Diabetic Mellitus. In their research, they classify diabetic results into 3 classes that is (1) Normal (2) Initial Stage (3) final stage. They have used map reduce technique that allows expressing distributed computations on huge amount of data and then they have used decision tree algorithm. It depends on yes or no environment based on diabetes results, where each non-leaf node denotes an investigation taking place a characteristic, each division characterize a conclusion of the investigation along with each side node holds a category label.

[15] Priyaga, Naveen et.al, proposed a web analytics support system using Naïve Bayes weighted approach that will gives prediction of the heart diseases. The dataset was taken from UCI repository. The dataset totally contains 76 attributes of 303 records, but their research concentrates only on 13 different attributes. The 13 attributes are as follows, patient age, patient sex, chest pain types, blood sugar level during fasting, ECG result, Exercise induced angina, slope, Thallium test, blood pressure of patient, cholesterol level, Thatch, old peak, CA.

Prediction of heart diseases will give the result among one of five classes. If the predictive system gives the value 0 then it indicates no heart diseases. If the value is 1 then it indicates less chance of heartdiseases. If the value is 2 it indicates average chance of heart diseases. If the value is 3 it indicates high chance of heart. If the value is 4 it indicates very high chance of heart diseases.
In their research, they have used Naïve Bayes algorithm for classification and data values in the dataset and patient input data are converted into weights. So that it will reduces the number of comparisons between the user values and the records of the dataset. In their research, the accuracy achieved is 86%.

Beside the research work on CRM, several companies are working on CRM. These companies are providing information on different phases of healthcare in their own way. A top-down strategy is mainly used for flow of information.

III. COMPARISON MODEL TO IDENTIFY THE ACCURACY IN ORDER TO PREDICT CANCER DISEASES USING DIFFERENT ALGORITHMS

Sl.No.	Cancer Type	Method	Number of Patients	Type of Data	Accuracy	Validation Method
1	Breast Cancer	ANN	62,219	Mammographic, Demographic	AUC=0.92(92%)	10 Fold Cross Validation
2	Multiple Myeloma	SVM	80	SNP'S	71%	Leave-one-out Cross Validation
3	Breast Cancer	SVN	174	SNP'S	69%	20 fold cross Validation
4	Colon Carcinomatosis	BN	53	Clinical, Pathologic	AUC=0.71(71%)	Cross Validation
5	Breast Cancer	SVM-CNN	650,356	Mammographic, Demographic	94%	10 fold cross Validation

Table : comparison of predictive model accuracy against other machine learning methods.

All new approach needs to focus on its implantation as well as the way of adaptation. This paper highlights on a accuracy identified by using various machine learning algorithms for the prediction of cancer disease.

IV. CONCLUSION

The Customer Relationship Management needs a customer-focused business strategy and background to support effective marketing sales and service process. Every service industry adopts CRM strategy to face large competition. The health care industry is not exempted as there is huge competition among private hospitals and a powerful CRM strategy is necessary for their survival. In coming days, hospitals will change their prototype from service-centric to patient-centric. The future of a hospital depends on the facility it offers to its patients. It should be more than the patient's expectations. Patients do not like questions with options for the services they would like to get offered to. The strength of a hospital largely depends on how they maintain contact with its patients. Even if the hospitals have premium class infrastructures, or high persona, they've got to maintain relationship with the patients as this helps the hospitals to retain their patients. There are different types of customer relationships, namely, Relationship between hospital personnel and patients, Relationship among different levels of hospital personnel and Relationship between Management and Employees. However, the aim of the case study is how CRM effectively works in hospitals. CRM in hospitals is a method of doing business to maintain patients for a long run. An attempt of such research was done in Madurai. This was done to suggest ways to improve CRM in corporate hospitals.

Customer Relationship Management is the keyword used in the present highly competitive and continuously changing business arena. CRM is a method used to learn more about consumers of their needs and their behaviors in order to build healthier relationships with the consumers. All matters at last is the bonds the companies maintain with the consumers, which in turn becomes the key of a business to be successful. Organizations say that all business processes from customer service to income have effects on the end user. CRM helps the customers enjoy following features:

(1)Customer Retention and loyalty.
(2)Rapid Return on Investment.
(3)Reduction in Sales cycle.
(4)Post-Implementation Support.
(5)Integration with other systems.

Nowadays, it has been a part of business corporates to change their focus to Customer Relationship Management, especially the private hospitals. Meanwhile some questions like these arise:

Current status of privately owned hospitals CRM-relation with healthcare providers Correlation between what patients expect and those of healthcare providers

In what way each healthcare provider presents its features to correlate with patient satisfaction on an overall basis In which all ways the overall patient satisfaction reflects the service quality singularly and collectively Perceptions of improvement of patient's satisfaction.

V. REFERENCES

[1] M.C.S Geetha, Dr.I.Elizabeth Shanthi, N.Sanfia Sehnaz, "Analyzing the Suitability of Relevant Classification Techniques On Medical Data Set For Better Prediction", IoT in Social, Mobile, Analytics and Cloud,2017,pp 665-670.

[2]Gunarathne.H.S.D,Perera K.D.M,Kahan dadawaarachchi K.A.D.C.P, "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease", computer society on Bioinformatics and Boengineering,2017,pp291-296.

[3]Tejaswini U Mane, "Smart Heart diseases Prediction System Using improved K-Means and ID3 on Big Data", International Conference on Data Management, analytics and Innovation, 2015pp239-245.

[4] Amrutha Valli P,Pravalika, Uma M.Sasikala T, "Tracing out Various Diseases by Analyzing Twitter Data Applying Data Mininig Techniques", Emerging Trends in Energy, Communication, Data Analytics and Soft Computing, 2017, pp1589-1593.

[5] Yuksel A.S,Cankaya S.F, Uncu I.S, "Design of a Machine Learning Based Predictive Analytics Systems for Spam Problem", Special Issue on Computational and Experimental Science and Engineering,2016,pp500-504.

[6]Suresh Kumar P, Pranavi S, "Performance Analysis of Medical Learning Algorithms on Diabetes Dataset Using Big Data Analytics", volume 9IEEE Access, 2017, pp508-513.

[7] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, Lin Wang, "Disease Prediction by Machine Learning over Big Data from Health Care Communities", Special Session on Health Care Big Data,2016,pp8869-8879.

[8] Rashmi G Soboji, Prem Kumar Ramesh, "A Scalable Solution for Heart Disease Prediction using Classification Mining Technique", International Conference on Energy, Communication, Data Analytics and Soft Computing,2017,pp1780-1785.

[9] Sonu S.R.Ravi Ranjan, Vivek Prakash, Saritha.K, "Prediction of Parkinson's Diseases using Data Mining" ,Emerging Trends in Energy, Communication, Data Analytics and Soft Computing, 2016, pp1082-1085.

[10] Alexandra L Heureux, Katarina Grolinger, "Machine Learning with Big Data: Challenges and Approaches", volume 5 IEEE Access, 2016, pp7776-7797.

[11] Milos Jordan ski, Milos Radovic, Zarko Milosevic, "Machine Learning approach for predicting Wall Shear Distribution for Abdominal Aortic Aneurysm and carotid Bifurcation Models ", IEEE Journal Of Biomedical and Health Informatics volume 22,2017,pp2168-2194.

[12] Shobana V, Kumar N, "A Personalized Recommendation Engine for Prediction of Disorders Using Big Data Analytics "Innovations in Green Energy and Healthcare Technologies, 2017, pp324-333.

[13] Peter K Ghavami, Kailash C .Kapur, "The Application of Multi Model Ensemble Approach as a Prognostic Method to Predict Patient Health Status", Chemical Engineering Transactions, vol 33,2013 , pp1056-1065.

[14] Thanga Prasad S, Sangavi S, Deepa A, Sariabanu, "Diabetic Data Analysis In Bigdata With Predictive Method", Computer and Electronics in Big Data, 2015, pp48-54.

[15] Priyanga, Naveen, "Web Analytics Support for Prediction Of Heart Diseases Using Naïve Bayes Weighted Approach", Asia modelling Symposium, 2017, pp21-26.