Application Tool for Blood Deferral Donor Model using Machine Learning Approaches

Abdul Rehman S.M[1],Shri Harsha C H[2], Nikhil M.G[3], Vandana Sardar [4]

[1,2,3]Student, Dept. of Computer Science and Engg., M.S Ramaiah Institute of Technology, India.
[4]Assistant Professor, Dept. of Computer Science and Engg., M.S Ramaiah Institute of Technology, Bangalore, India.

Abstract

Blood is a vital thing for living being because there is no alternative replacement for it, Hence it makes it necessary for a development of a system which keep track of the healthy blood by predicting continuous behavior of blood donors in blood transfusion organization. The main aim of predicting the behavior pattern of donor and blood is to predict whether the donor is eligible to donor blood. To do so we gather some valuable information from the donor and based upon the information the prediction is done whether the donor is Deferral or Non-Deferral and donor data is stored in the Database to enhance the performance donor selection process in the transfusion.

Keywords— *Donor, Deferral, Prédiction, Transfusion, performance (key words)*

I. INTRODUCTION

Blood is always an important entity in human being as well as all the living beings, because it plays a vital role in human body an average human has 4.5 to 5.5 litres of blood. Blood is always in great demand from past, nowadays in increase in population growth we need blood in several aspects due to accidents, Surgeries ,Diseases etc. Blood donation and Transfusion is an indispensable part of medicine. Nowadays information and technology has reaching new heights as we have advanced technology for several things. Hence we need a model which will be helpful in transfusion of blood. These days technology has made many remarkable development in field of medicine or biology such as new techniques of surgery, scanning etc. and when it comes to blood donation and transfusion still we are stuck on the same criteria using the database approach, though database approach has many pros even it has some cons such as 1. Maintaining currency 2.Management Complexity. 3.Needs frequent upgrade etc. hence the intended model is based upon Machine Learning technique which uses several aspects to make blood transfusion and donor data more relevant.

This model basically aims on classification of blood donors into two types Deferral and Non-Deferral donors. Deferral are those donors who are unfit to donate blood and non-deferral are those donors who can donate the blood without any restrictions. Initially when a donor arrives to donate blood several questionnaire are done based upon their medical health based upon the information provided we use Machine Learning approach to classify the donor. The classification uses several criterion and algorithm and process on a dataset based upon which the results are provided. Initially a model is trained by providing small dataset to evaluate the accuracy of prediction later we check how well the model works on large dataset and calculate its efficiency for the given dataset. Here we use K-NN algorithm in machine learning approach which works upon finding its nearest neighbour, the information which is present in a dataset may vary from one entity to another hence we use the K-NN Algorithm which will efficiently work in distinguishing the donor type.

II. Related Work

Blood may be a deliverance drugs for critically sick patients, because it saves several lives. So, within the blood donation. it's vital to take care of safety of donor and recipient. For that donor choice is most significant criterion of blood donation. Donor choice is denied because the method of assessing the suitableness of a private to give blood or blood elements against totally different choice criteria. correct and correct donor choice will avoid transfusion reaction. the protection and availableness of blood and blood merchandise for transfusion needs the accomplishment and choice of voluntary non-remunerated blood donors. The safest blood

donors ar voluntary non-remunerated blood donors from low risk population. we have a tendency to develop a model of characteristic deferral donors and reason for deferral.

According to transfusion service in Bangla Desh , the safe transfusion programme, presently quite vi,00,000 units of blood is needed annually in Bangla Desh once we take into account this case in inhabited Countries like Bharat and China the necessity could go terribly high, As per survey on thirty first of the demand is consummated from voluntary blood donations, whereas the remainder comes from relatives and or replacements donors . thence this model is meant to stay track of its healthy donors and motivating them to give blood once regular intervals . the donors ar classified temporary deferral and permanent deferral . Permanent deferral cannot give blood in there entire life however, whereas the temporary deferral ar caused due some medical problems like ColdFlu, Malaria, Jaundice etc. once the patient approaches for donation thanks to there improper recovery we have a tendency to could result into loss of a valuable donor. thence this model keep track of all the donor knowledge and classifies the donor whether or not he or she may be a temporary or a permanent donor. If the donor is temporary then it's to supply its next donation date to voluntarily give the blood. thence we have a tendency to use many Machine Learning approaches to classify and make out the causes for deferral in order that the donor will retaliate his health.

The K-NN rule uses the closest neighbor approach to search out its corresponding entity. This rule stores all the offered cases and classifies new cases primarily based upon the similarity measures. it's essentially utilized in estimation and pattern recognition.

A case is assessed by a majority of vote of its neighbor, with the case of being appointed to commonest among the k nearest neighbors is measured by the space perform referred to as K. easy if K=1 then it's appointed to its nearest neighbor. It takes 3 functions to search out the space one.Euclidean a pair of.Manhattan three.Minkowski. it ought to be noted that every one the 3 distance perform measures ar valid just for continuous variables. within the instance of categorical variables the performing distances should be used. It additionally brings up the difficulty of standardization of the numerical variables between zero and one once there's a mixture of numerical and categorical variables ar gift within the dataset. as a result of the machine learning rule ar wide accustomed render potential resolution to most the issues. several rule ar enforced and compared the performance with relevancy given resolution. Most of them belong to medical field like call Trees(DT), Support Vector Machine(SVM),Artificial Neural Networks (ANN) and Random Forest(RF) etc. these ar all the rule that play associate important role in drugs.

III. EXISTING SYSTEM

Since this problem is classical , there are several methods of handling it. It can be classified into three categories.
1.Physical Interview 2. Centrally Digitalized or DBMS approach.
3. Machine Learning (gradual adoption happening).

1.This method was widely used in earlier days where blood was collected from the donors and its is later taken to blood bank and tested. Where there was no proper information of donor and blood type still it is being practiced in few countries. 2. The DBMS approach this system had a widespread of usage this system takes all the information of the donor and blood types and other relevant necessary information and is stored in a database. These system has several advantages and disadvantages as well.
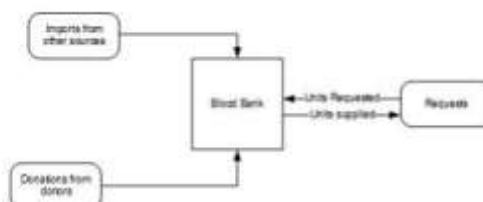

Figure 1 General Model of a Blood Bank.

Disadvantages of Previous Approach

The main disadvantages of the DB system and Physical Interview approach is Increased cost ,More Time Consumption ,Irrelevant data, Management Complexity, Frequent Update and replacement cycles etc.

IV.   **PROPOSED SYSTEM**

Methodology

We present Machine Learning approach where Data is collected from different donors , the donors are evaluated on the basis of screening criteria which involves questionnaires having personal details, Clinical details and Physical Examination Such as Blood Pressure, Hemoglobin Estimation and Body temperature etc.
It mainly categorize the donors into two type deferral and non-deferral donors . the guidelines were used for deferral blood donors . Deferral donors data is analyzed with respect to age, gender ,type of donor and causes for deferral were categorized into temporary and permanent, based on the curability of the condition using machine learning algorithms .We are here using SVM ,K-NN and K-means algorithm for classification and prediction. Later , will compare them based on their efficiency to predict deferrals. In second part if they are deferral , firstly reason of deferral is found out then we suggest them diet tips or health check-up advice according to the reason (for example ,if a person has low hemoglobin count we would suggest him to take up food which increase the count of hemoglobin such as beetroot) we will send them SMS or email and conjointly inform them to come back back on bound date when their recovery.
From this we have a tendency to shed the sunshine on the health of people which {can|which is able to} produce awareness among them and that we can avoid loosing voluntary non-remunerate donors. Donor choice is outlined because the method of assessing the quality of a personal to give blood or blood elements against totally different choice criteria. correct and correct donor choice will avoid blood response conjointly.   The safety and convenience of blood and blood product for transfusion needs the enlisting and choice of voluntary non- salaried blood donors. The safest blood donors area unit voluntary non-remunerated blood donors from low-risk population. they're unpaid donors. Blood transfusion service ought to aim at eliminating the risks and hazards of transfusion. Donor quality is most necessary. Donors ought to be questioned regarding their case history and ought to be given a mini-physical examination to assist blood centre workers decide whether or not theyarea unit appropriate donors.

Fig.1: A simplified illustration of the blood supply-chain system

V.**SYSTEM ANALYSIS AND DESIGN**

Donor selection is defined as the process of assessing the suitability of an individual to donate blood. Initially the model collects the all the relevant information as discussed in [3].The questionnaire will be based upon the several aspects of the donors and based on the data provided the data is processed with the trained data set . after the process the system will distinguish whether the donor is deferred or non-deferred depending on the donor types necessary actions are taken. And the system also classifies the donors as deferred and non-deferred , if the patient is not deferred then the donor quality is evaluated and then it provides the date to the next donation dates.

System model

Machine learning algorithms are widely used to render potential solutions to real-life problems in the field of computer science. Many algorithms have been implemented and compared for performance and efficiency by employing them on various datasets, most of them belonging to medical fields. Decision Trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Bayesian Networks, Support Vector Machines (SVM) and Gaussian Processes (GP) are among the most frequently used. A comparative study on several of these methods indicates what algorithms suit a given data set the best. A dataset of ICU patients with head injuries, classified using DT reveals that the implementation is not only easily comprehendible but also has a moderately high resistance to noise and errors. It basically use 4 rules.

1.1 CLASSIFICATION

Classification is the predefined groups or classes. Classification is that the estimation information|the info|the information} and it's the algorithms square measure requiring to the category and outline to data attribute values. The thrombocyte transfusion is mistreatment the classification techniques square measure rule induction, call tree, neural network and back propagation.

1.2 CLUSTERING

Clustering techniques are using the algorithm are K Nearest neighbour means algorithm, and k medoid algorithm. Clustering is the grouping the data item.

1.3 ASSOCIATION RULE

Association rule is created and analysing the data .the if /then patterns are used to the two relationship . There are support and confidence .support is how the data items are frequently seem in database. Confidence is how many times the if/then statement

1.4 REGRESSION

Regression is that the perform function mathematical relation} it's notice a function and that the models of information with least error.Rule – induction within the rule induction is that the exploitation if/then patterns .The rules are used in two parts .There are antecedent (if part)and consequent (then part). Example of the thrombocyte transfusion of no-hit or unsuccessful in IF-THEN Rule induction.
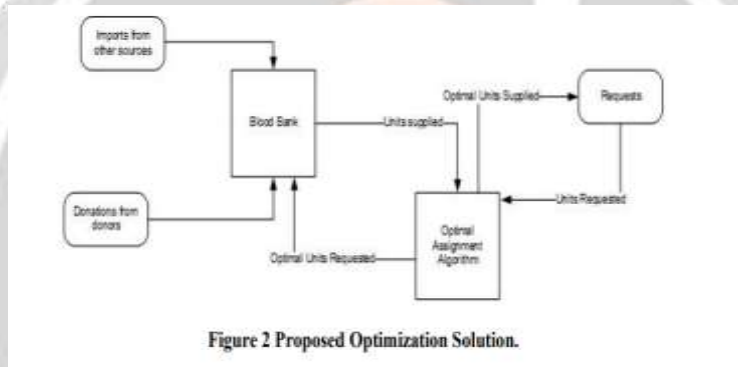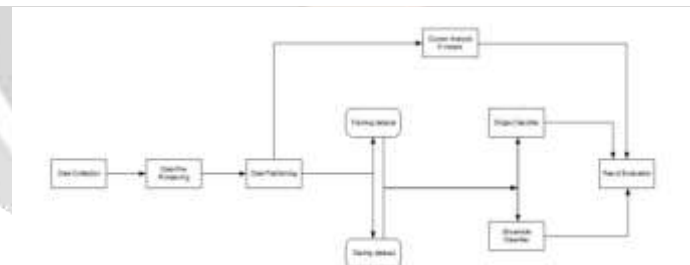

Figure 2 Proposed Optimization Solution.

Figure 3 Proposed System Flow.

Algorithm

Inputs and initializations:

K-nearest neighbour (K-nn) algorithm :

Let (Xi, Ci) where i = 1, 2……., n be data points. Xi denotes feature values & Ci denotes labels for Xi foreach i.

Assuming the number of classes as 'c'

Ci ∈ {1, 2, 3, ……, c} for all values of i

Let x be a point for which label is not known, and we would like to find the label class using k-nearest neighbour algorithms.

Calculate "d(x, xi)" i =1, 2, ..., n; where d denotes the Euclidean distance between the points.

Arrange the calculated n Euclidean distances in non-decreasing order.

Let k be a +ve integer, take the first k distances from this sorted list.

Find those k-points corresponding to these k-distances.

Let ki denotes the number of points belonging to the ith class among k points i.e. k ≥ 0

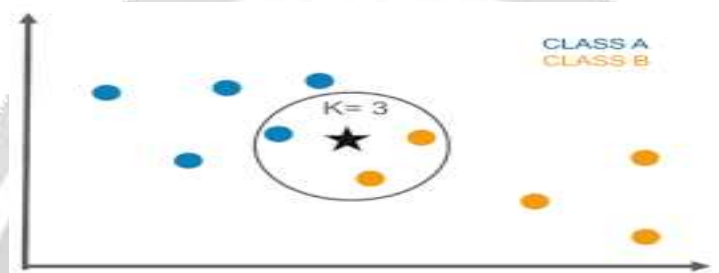**If ki >kj ∀ i ≠ j then put x in class i.**

End

Fig 5.1 & 5.2 Representation of Parameter K in K-NN Algorithm

Work load

The K-NN Algorithm has good efficient in processing of huge dataset and the algorithm also works well for the robust and noisy data and it is effective if the training data is large and it need to determine the value of parameter K

Results and Discussion

The result based upon the given dataset works fine if the training data is large and the parameter K should be defined as well. And the computation cost is quite high because we need to compute the distance of each query instance to all training samples.

VI Expected Results and Future Scope

As the size of the dataset increase it will be easy for the model to analyze and we can obtain maximum accuracy in prediction of the results. We can use several other machine learning approaches such as SVM , K-means clustering and other neural network approaches to make the model to work more efficiently

Classifier	Training Set	Test Set	Accuracy rate (in %)
SVM	10,000	10,000	98.9
KNN	10,000	10,000	96.47

TABLE 1: Accuracy of Classifiers

VI. CONCLUSION

By referring some of the journals and papers related to this criteria, the motivation for this project concerned for the donors as well as the faculties to distinguish the deferral and non-deferral donors. In order to easily classify deferred and non- deferred donors based on the algorithm the data is evaluated to achieve maximum no of non-deferral donors .So that they can donate blood . This project is solely being developed to achieve the quality blood for transfusion and it mainly aims at eliminating the risks and hazards of blood transfusion. a robust model was constructed for the purpose set out. The data was extracted from 5 major listed firms on the Jordanian stock market, the sample information was wont to be our coaching information set (about two hundred records for each company) upon the factors antecedently mentioned to use our model. We adopted associate degree economical prediction algorithmic program tool of kNN with k=5 to perform such tests on the coaching knowledge sets we have a tendency to had. According to the results, k-NN algorithm was stable and robust with small error ratio, so the results were rational and reasonable this presents a good indication that the use of knowledge mining techniques might facilitate call manufacturers at numerous levels once exploitation kNN for data analysis. So, we consider that employing this prediction model, kNN is real and viable for prediction of the donor type.

REFERENCES

[1] NetrikaW. Dhoke, Prof. Swati S. Deshmukh Student of M.E, Dept of CSE, Raisoni College of Engineering and Management, Amravati, India. To Impore Blood Donation process using Data Mining Techniques IJIRCCE Vol.3,Issue 5,May 2015.

[2] Arunkumar Chinnaswamy,Gurusankar gopikrishnan,Kamal Kiran Pandala ,Shabala Natrajan. Dept of CSE Amrita School of Engg. Coimbatore,India Vol 10, No 7 (2015)

[3]Deepthi Srambical Poulose and Dinesh Kumar Sahu and Anil Rajput. Dept of CSE ,Sri Satya Sai College of Engg. Bhopal ,India An Experimental Study of Data Mining Techniques in Blood Donor Sector. Vol 4 Issue 5 May 2015

[4] L Bala Senthil Murugan, Anitha Julian, "Design and implementation of automated blood bank using embedded systems", Circuit Power and Computing Technologies (ICCPCT) 2015 International Conference on, pp. 1-6, 2015

[5] T. Santhanm and Shyam Sunderam: 'Application of Cart Algorithm in Blood donor's classification', Journal of computer Science Vol. 6, Issue 5.

[6] Shyam Sundaram and Santhanam T; 'Real Time Blood donor management using Dash boards based on Data mining models', International Journal of Computing issues, Vol.8, Issue5, No.2, September 2011

[7] Dr Amita Goel ,Saarthak Mahajan. Dept of ISE Maharaja Agrasen Institute of Technology. New Delhi ,India Comparison between K-NN and SVM . Vol 5 Issue 11 December 2017

[8] Arul Valan ,Dr.E. Babu Raj Professor , Dept of CSE Marian College Trivandrum, India. Vol-4 Issue 12 Dec 2018

[9] Nagurney, A., A. H. Masoumi, and M. Yu (2012). "Supply chain network operations management of a blood banking system with cost and risk minimization". In: Computational management science 9, pp. 205–231.

[10] Schreier, G. (2010). Pervasive healthcare via the Internet of Medical Things. Graz, Austria: Austrian Institute of Technology GmbH.