International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April, 2020

Medibot: A Predictive Generic Diabetic Chatbot using Bagging Ensemble/Hybrid Learning

Saritha A K

Abstract*: Clinical chatbots are conversational operators worked in light of clinical applications. They can possibly lessen medicinal services costs and improve availability of clinical information to basic man. There are different methods accessible for planning chatbots for anticipating an infection. In any case, a client can accomplish the genuine advantage of a chatbot just when he can connect with it in a simple manner and it ready to foresee the infection with high level of exactness while simultaneously give all important data being looked for by the patient. Chatbots can either be conventional or sickness explicit in nature. Diabetes is a non-infectious ceaseless human issue. Early forecast of this issue can uncover the deplorable intricacies and help to spare human life. Right now, have first built up a conventional book to-content 'Medibot' – a chatBOT which connects with patients in discussion utilizing propelled Natural Language Understanding (NLU) methods to give customized forecast dependent on the different side effects shared by the patient. The plan is additionally stretched out as a chatBOT to diagonise particular Diabetes type expectation and for proposing prevention measures to be taken. For expectation, there exists various grouping calculations in ML Ways which can be utilized dependent on their exactness. Nonetheless, as opposed to thinking about just one model and trusting this model is the best/most exact indicator we can make, the curiosity right now in Hybrid Algo realizing which is a meta-calculation that joins a bunch of models and midpoints them to create one last model to diminish change (stowing), predisposition (boosting), or improve expectations (stacking). From writing surveys, it is seen that almost no exertion has been placed into utilizing troupe techniques to expand expectation precision. The paper introduces a cutting edge Medibot plan with an undemanding front-end interface for normal man utilizing UI, NLU based content pre-preparing, quantitative execution examination of different AI calculations like Gaussian Naïve Bayes , Entropy Decision tree, Random Forest, K- NN, Support Vector Machines, Logistic and X-Gradient boosting as independent classifiers and joining them all in a dominant part casting a ballot troupe for adjusted outcomes. It is seen that the chatbot can interface consistently with any patient and dependent on the side effects shared, anticipate and rank the most likely ailment precisely utilizing the nonexclusive model and explicitly diabetes dependent on a strong outfit learning model.

Keywords* : *Bagging , Ensemble learning, Generic Dataset , Gradient boosting, K-nearest neighbour, Logistic Regression, Majority voting ensemble, Multi Layer perceptron, Naïve Bayes, Natural Language Processing, Neural Networks, Random Forest, Stacking, Voting.

Revised Manuscript Received on April 25, 2020.

* Correspondence Author

Saritha A K*, Dept of CSE, Gitam University, Bangalore, India. Email:aksaritha@gmail.com

I. INTRODUCTION

By applying computational examination on clinical huge information, the gigantic measure of information produced in the human services frameworks, will be utilized to make clinical insight which will drive clinical expectation and guaging. Clinical investigation is another pattern in clinical science. Creating clinical knowledge out of the clinical information accessible will make medicinal services framework to show restraint driven and will decrease clinical expense and emergency clinic readmission as well. A great deal of sicknesses can be restored whenever analyzed well ahead of time. A chatbot is a PC program that leads a discussion by means of printed or discourse component. Today Chatbots are developing at a pace what sites were in late 90's and 2010 as appeared in Fig 1. Clinical chatbots are such conversational operators worked in view of Medical applications like diagnosing a conventional ailment or for persistent consideration. Like in Medical speech, we have general doctor and a master, right now, mirror a comparative state utilizing NLU and Advanced ML calculations to initially analyze a malady utilizing a nonexclusive content to-content Constextualmedibot and afterward broaden this conclusion as a specialization to increasingly more profound level forecasts in the investigation of diabetes. Diabetes is a non-transferable consistent human disease. Early estimate of this illness can tell a patient of the authentic ensnarements and help to save human life. The odd augmentation in heaviness and stationary lifestyle has made diabetes a general scourge. The fundamental clarification for this disease is the overdose of glucose in the blood. Human body changes by far most of the sustenance into glucose, and it is the basic wellspring of essentialness in the body.. An individual becomes diabetic when his/her body can't change glucose into vitality. Because of unreasonable glucose in the blood, BSL (Blood Sugar Level) builds that at last prompts diabetes [1] [2]. Diabetes has risen as a significant social insurance issue in India. As per diabetes chart book distributed by the International diabetes alliance (IDF), there were an expected 50 million people with diabetes in India and this number is anticipated to ascend to just about 80 million individuals by 2025. It is assessed that each fifth individual with diabetes will be an Indian. As per the Health Organization in world, it was evaluated that 4.4 million passings are caused because of high glucose [3]. It has been indicated that early recognizable proof of type 2 diabetes can help forestall or postpone 80% of intricacies [2] and lessen the

quantity of passings brought about by the entanglements. Additionally diabetes is the first reason for visual impairment, removal and kidney disappointment. Along these lines, it is attractive to accurately recognize individuals in danger of diabetes in an auspiciousway.

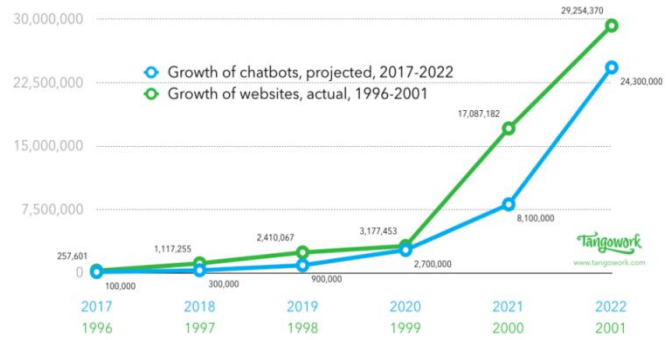


Fig 1: Growth of Chatbots

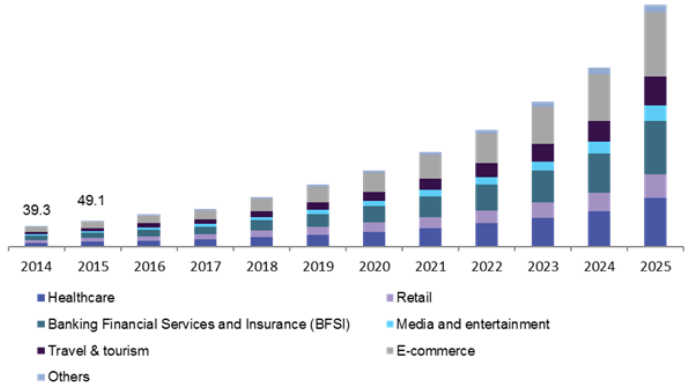
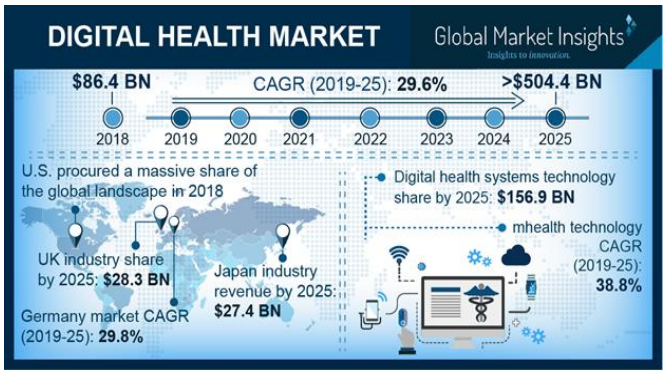


Fig 2 : Growth of Chatbots in Different Segments



Digital Market in Health Industry.

To analyze diabetes, a doctor needs to investigate patient's information and think about numerous variables (for example family ancestry, age, weight file, and so forth.). A doctor's conclusion can be emotional and is exceptionally reliant on his accessibility and encounters. right now, present a UI based front-end NLU chatbot which draws in with the patient and checks for different side effects. It predicts different infections conventionally. Whenever anticipated diabetes, it goes further into the malady to check type and different subtleties. We present a methodology that joins an outfit of five classifiers – Machine learning based 5 Algorithms and Neural Networks on account of nonexclusive illness forecast utilizing the General heath dataset. On account of Diabetes expectation, an outfit of 6 classifiers are applied on the Pima Indian diabetes (PID) informational index [13].

The rest of the areas of the introduced original copy are isolated as follows. Segment II talks about the related work, which is trailed by the conversation of the different existing classifiers for diabetes expectation. Segment IV examines the proposed framework and its execution. Area V presents the outcomes and conversation, while the general outcomes and end is introduced in Section VI. References utilized are given toward the finish of the composition.

II. PREVIOUS REVIEW

This segment briefs a portion of the key research done in chatbot plan and grouping for malady analysis.

A chatbot resembles a typical application. There is an application layer, a database and APIs to call other outside organizations. Clients can without much of a stretch access chatbots, it adds unpredictability for the application to deal with. Be that as it may, there is a typical issue that must be handled. It can't grasp the arrangement of the client. Right now, bots are prepared by the past trained data model accessible to them. In this way, most associations have a chatbot that keeps up logs of conversations. Data scientist use these logs to examine what customers are attempting to inquire. With a mix of AI instruments and models, Engineers facilitate customer requests and answer with the best proper answer. For instance, if any client is getting some information about installments and receipts, for example, "where is my item installment receipt?" and "I haven't got an installment receipt?", the two sentences are taken to have a similar importance.

In the event that there is no far reaching information accessible, at that point diverse APIs can be used to prepare the chatbot.

Examiners have so far worked in isolation with aversion to uncover any improved frameworks that they have found, impeding the moves up to chatbot structure. Also generally valuable chatbots need upgrades by organizing dynamically complete data bases which are insufficient with respect to today. N. Jyothirmayi et all [21] found that chatbots need mining result as needed. Moreover, they don't contain satisfactory basic language traits. Monica Agrawal et all in their investigation endeavored to collect a book to-content examination bot that associates with patients in conversation about their clinical issues and gives a tweaked discovering reliant on their reactions and profile. Notwithstanding, their algorithmic exhibition experienced low accuracy, review and anticipated analysis rates. Additionally they proposed including support for progressively clinical highlights, for example, area, length, force of side effects, increasingly point by point indication depiction, provoking and recommender framework and so on to improve execution of the chatbot.

Following highlights are normally assessed well and patients accept that seeing this information improves health. However, these evaluations and convictions about adequacy ought not be mistaken for genuine drivers of conduct change. While a few types of self-checking, for example, recording diet and exercise, can be compelling in certain cases,22 rising proof recommends that demonstrating a patient patterns in their clinical qualities isn't successful at changing behavior.

Along these lines, there is a requirement for better algorithmic arrangement execution for finding by means of option and developing registering procedures. For chatbot plan, an assortment of strategies have been proposed yet there exists no speculation. There is a requirement for a simple UI with adequate Natural language understanding which a typical man can utilize. Likewise, very little research has been completed to assemble a solitary framework joining an undemanding UI-based chatbot, a conventional though exact beginning ailment expectation and as far as possible up to a specific diabetes forecast utilizing propelled AI procedures, for example, group learning.

III. MY WORK

The commitments of this examination are cumbersome and conplex. A bit of the key commitments of the displayed look into work are given as follows:

Advancement of an easy to use content to-content talk interface at the front-end to connect with the patient Nonexclusive illness expectation classifiers with quantitative execution appraisals alongside a group or meta-classifier consolidating all classifiers

Diabetes expectation classifiers with quantitative execution evaluations alongside a gathering or meta-classifier joining all classifiers

Contrasted and different looks into, the methodology offers a start to finish Medical chatbot structure with a simple UI, conventional forecast and a malady explicit specialization execution. With improved exactness utilizing group learning, more diabetes patients can get in-time data and take proper activities, for example, evolving way of life to forestall or postpone the inconveniences of this ailment.

IV. ARCHITECTURE OF MEDIBOT

At the broad end my architecture consists of React Framework based UI/ Inteface for user to Interact with the bot.React Framework is given by facebook and Javascript.TheMedi-bot conversates with Natural Language Understanding Engine at the backend loaded with necessary libraries from RASA NLU.NLU is trained with diabetes dataset collected from UCI repository.ThisMedi bot helps in preliminary check NLU engine which is set up to give yield subject to customer requests.

A brief schematic of the structure is showed up in Below diagram.

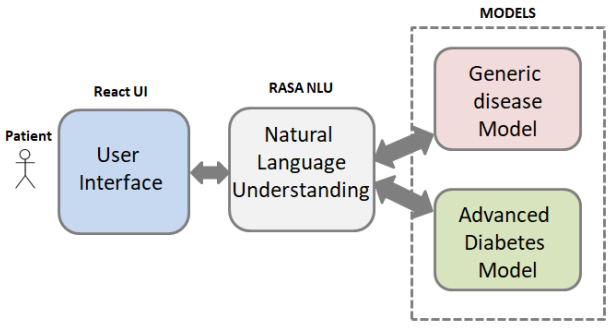


Fig 3a: Medibot System architecture

In Depth Architecture of Model training,tested and validated with Voting,bagging and Boosting Hybrid Model is presented in detail.

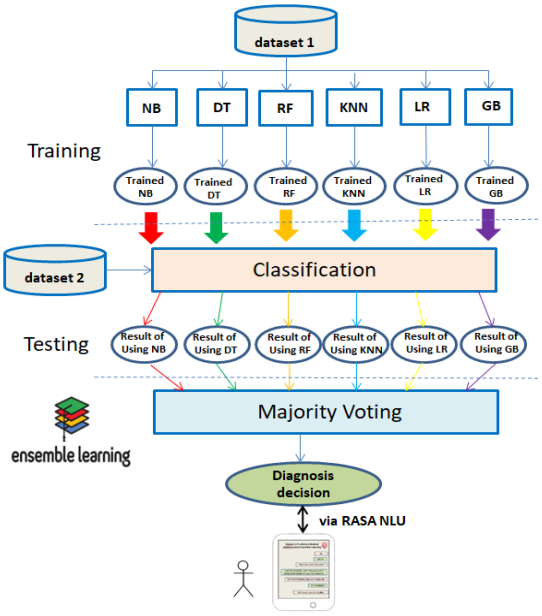


Fig 3b: Hybrid Learning Testing and validation

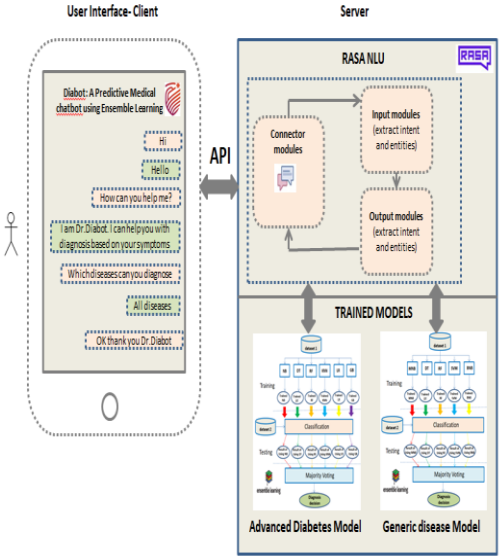


Fig 4: Detailed Medibot architecture

The point by point designing of the Medibot is showed up in Fig 4. The Architecture is based on low latency communication server based development where UI is made using the React Framework written in Javascript which is opensource given by facebook.. It talks with a readied RASA engine at the backend. The NLU engine is related with both the regular prosperity and the impelled diabetes model and their yields (Model data) are saved in Pickle record gathering. On tolerating an API can, the model data in the pickle record configuration is stacked and made available to the UI in Flask server (webserver) gathering. The genuine picture of the chatbot UI made is showed up in Fig 5.

Hybrid Learning

One of the key commitments right now Hybrid learning.

Half and half Model relies upon the assumption that different models arranged unreservedly are likely going to be valuable for different reasons: each model looks possibly changed pieces of data to make figures, getting some part of the real world yet only one out of every odd last piece of it [25]. The notable techniques for merging the classifiers in troupe learning are mix of masters, larger part throwing a voting form gathering, boosting, stowing and stacking. Lion's offer throwing a voting form bunch is actually a combiner methodology that can be used close by stacking based assembling learning. Stacking relies upon a heterogeneous game plan of feeble understudies. Every classifier is readied self-rulingly

what's more, last decision is made by a lion's share vote, averaging the outcome [26]. Current work utilizes stacking-based group learning with dominant part casting a ballot gathering as combiner. The frail students are Support Vector Machine(SVM),Random forest(RF)Entropy Gain Decision tree(DT), Decision based Random Forest(RF), , and Bernoulli Naïve Bayes(BNB),Gaussian Naïve Bayes(GNB), if there should be an occurrence of conventional illness model and Naïve Bayes(NB), Decision tree(DT), , K-closest neighbour(KNN), Logistic regression(LR) and XG boosting(GB) on account of model

The dataset is part into 70:30 proportions for preparing and testing the exhibition of every classifier. To join the presentation of every classifier, the accompanying larger part outfit casting a ballot rule is utilized. It takes a shot at names. Beneath dt,j can be 1 or 0. On the off chance that classifier 't' picks 'j', the worth is 1. At that point the troupe picks class 'J' which gets the biggest complete number of votes:

$$\sum_{t=1}^{T} d_{t,j}(X) = max_{j=1,\ldots c} \sum_{t=1}^{T} d_{t,j}$$

For nonexclusive infection figure, the general prosperity dataset is considered. It is a table with various signs/features and a related affliction for a ton of above appearances. In taking care of, we have made a transpose of this cross section and consigned characteristics to all features. There are around 40 highlights and have been used in getting ready and testing our counts.

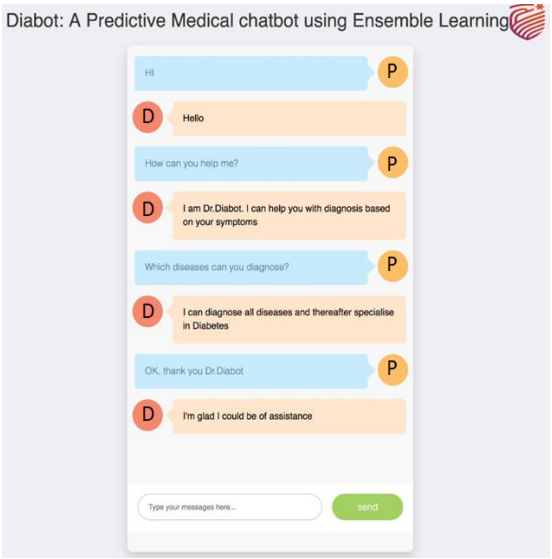


Fig 5: Genuine picture of the User Interface created

For diabetes gauge, we used the Pima Indian diabetes instructive assortment, one of the generally used datasets for desire for type 2 diabetes. It is from the University of California, Irvine (UCI) AI file [19].The educational assortment has 268 diabetes patients and 500 regular subjects. All subjects are females who are in any occasion 21 years old and of Pima Indian inheritance. The models involve 8 attributes and one pointer/class variable which have values 0 and 1. 1 shows tried positive for diabetes (demonstrated by yield one) or not (showed by zero).

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

We do information pre-preparing and evacuate for instance esteems like zero which are naturally incomprehensible (for example pulse). Next, highlight determination is done and we discovered that all characteristics are required as there exists solid relationship among them and for better exactness as appeared in Fig 6. It shows that darker the shading, the more reliance between the factors exists. We additionally plot a histogram of probabilities between every one of the trait being a 0 or 1 as appeared in Fig 7.

05 frail students or calculations if there should arise an occurrence of nonexclusive sickness model and 06 powerless students or calculations in the event of diabetes model are run on the information to gauge precision. Gathering model consolidates the presentation of every single frail student and precision estimated.

V. RESULTS AND METRICS

The framework of relationships between's the factors is appeared in Fig 6 and the histogram of probabilities between every one of the properties being 0 or 1 is shown in Fig 7.

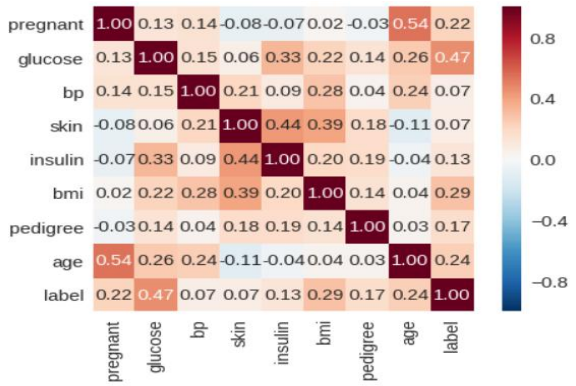


Fig 6: Matrix of correlation between variable

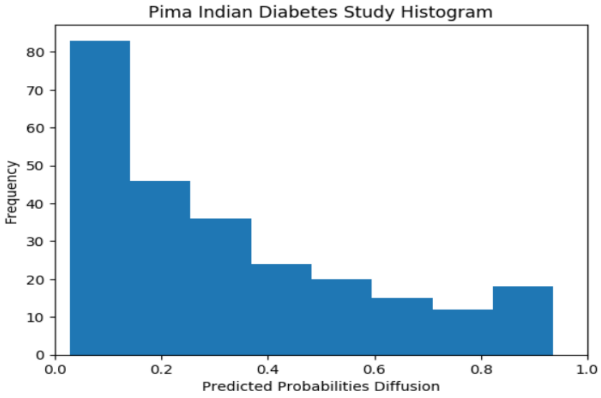


Fig 7: Graphical study of probabilities between attributes

Table 1 presents the presentation of each single classifier and the outfit of classifier utilizing dominant part deciding in favor of the conventional malady dataset. compared to Ensemble Gradient Boosting and Random Forest have performed good.But we are considering all possible test cases better in Ensemble –Voting Classifier. Fig 8 analyzes the exactness of each of the 05 models. The Decision Tree model has the most reduced exactness. Anyway the distinction of exactness between these 05 classifiers isn't huge. Table 2 presents the presentation of each single classifier and the troupe of classifier utilizing larger part deciding in favor of the Pima Indian diabetes dataset. The exhibition of the hybrid classifier is higher than every single other classifier.

Table 1: Precision Inference of Algorithm(s) for health dataset

	Algorithm	Precision (%)
i	Gaussian NB	90.89
ii	Entropy Random Forest	84.96
iii	Entropy Gain Decision Tree	80.76
iV	Support Vector Machine	82.89
V	Bernoulli -NB	84.58
VI	**Bagging-Voting Ensemble classifier**	**87**

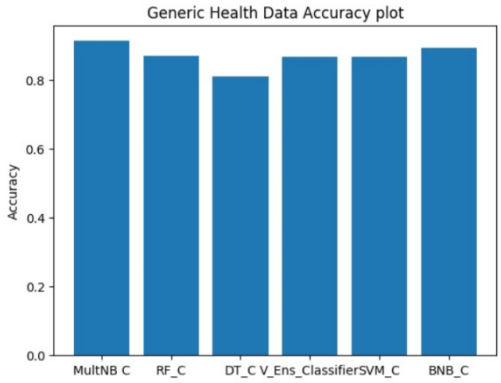


Fig 8: inference of Model accuracy based on Health data

Fig 9 compares the accuracy of all 6 models. The Naïve Bayes and Decision Tree model have the lowest accuracy. However the difference of accuracy between Random Forest and Ensemble learning is very near.

Table 2: Precision comparison of Models(s) for Diabetes dataset

.	Algorithm Name	Precision(%)
a	Gaussian NB	61
b	Entropy -Decision tree	65.2
c	RF-Random Forest	79.98
d	KNN	74.66
e	LR-Logistic Regressor	86
f	X-Gradient Boost	82.68
g	**Bagging-Voting**	85.0

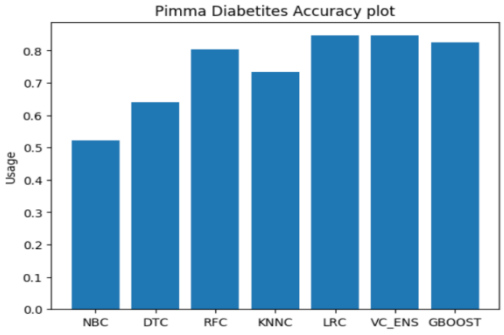


Fig 9: inference of Model Precision (Diabetes data)

Below Plot compares all the ROC curve which good and validated method which improved the accuracy.

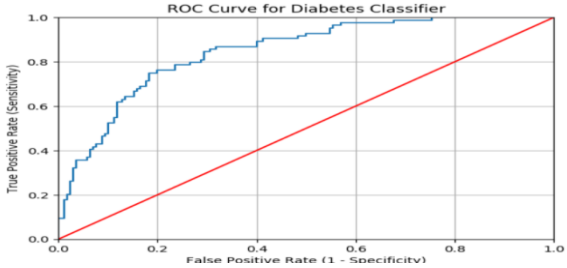


Fig 10: Diabetes classifier Receiver operating curve

VI. CONCLUSION AND FUTURESCOPE

To add up to, our through and through methodology for assurance of a regular disorder and subsequently go further into any one disease like diabetes has various focal points. At first, the regular framework suggested can be used as an autonomous model for any wide prosperity conjecture. This will assist slash with bringing affordability and license of users to focuses or facilities for basic testing and further medication. Further, the nonexclusive structure can be connected with developing any contamination unequivocal clinical chatbot. This we have demonstrated again by methods for the Medical diabetes chatbot, which is a specific utilization for those contamination unequivocal features. This will help predict critical sufferings early for proactive preventive measures and any lifestyle changes. Since the decision is made on a collecting all models instead of a single model which may manage, it is a clear and capable system while giving extraordinary execution. Our future work will concentrate on pushing important data genuinely into a medibot server using bluetooth and building a processing framework using the nonexclusive structure prescribed here.

Since PDAs are getting stunning ceaselessly and are seen as near and dear partners for their user's, fusing the prosperity parameters continuously like health data from wearable's, diet, water utilization, etc can be used by methods for API joining onto our organized medibot application enhancement for Android and iOS versatile stage.At the end of the day, we will recollect more models for the Development for diagnosing diabetes in all patients. Moreover, we will remember the classifiers with improved execution for the gathering. Additionally, the procedure will be applied to break down various diseases.

REFERENCES

1 An overview of artificial intelligence based chatbots and an example chatbot applicationNazAlbayrak ; AydenizÖzdemir ;EnginZeydan 2018 26th Signal Processing and Communications Applications Conference (SIU).
2 Enterprise Crowd Computing for Human Aided Chatbots Alessandro Bozzon 2018 IEEE/ACM 1st International Workshopon Software Engineering for Cognitive Services.
3 Intelligent Chatbot for Easy Web-Analytics Insights Ramya Ravi2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
4 Chinese sentence based lexical similarity measure for artificial intelligence chatbot Wen Zhang ; Heng Wang ; Kaijun Ren ; Junqiang Song 2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI).
5 AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT) Neelkumar P. Patel ; Devangi R. Parikh ; Darshan A. Patel ; Ronak R. Patel 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA).
6 Automated Thai-FAQ Chatbot using RNN-LSTM PanitanMuangkammuen ; Narong Intiruk ; Kanda RunapongsaSaikaew 2018 22nd International Computer Science and Engineering Conference (ICSEC).
7 Artificial Intelligence Marketing: Chatbots UrošArsenijevic ; MarijaJovic 2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI).
8 Comprehensive Technology Function Product Matrix for Intelligent Chatbot Patent Mining N.J. Oscar Hong ; UsharaniHareesh Govindarajan ; Y.C. Jack Chang Chien ; J.C. Amy Trappey 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC).
9 Expression Tracking with OpenCV Deep Learning for a Development of Emotionally Aware Chatbots Karmelo Antonio Lazaro R. Carranza ; Joshua Manalili ; Nilo T. Bugtai ; Renann G. Baldovino 2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA).
10 A Pilot Study Integrating an AI-driven Chatbot in an Introductory Programming Course Matthew Verleger ; James Pembridge 2018 IEEE Frontiers in Education Conference (FIE).
11 Lead Engagement by Automated Real Estate Chatbot Tho Quan ; Trung Trinh ; Dang Ngo ; Hon Pham ; Long Hoang ; Hung Hoang ; Thanh Thai ; Phong Vo ; Dang Pham ; Trung Mai 2018 5th NAFOSTED Conference on Information and Computer Science (NICS).
12 A Platform for Human-Chatbot Interaction Using Python Bhaumik Kohli ; Tanupriya Choudhury ; Shilpi Sharma ; Praveen Kumar 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT).
13 Pima Indians Diabetes Data Set, [Online], Available: http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes
14 Programming challenges of chatbot: Current and future prospectiveA M Rahman ; Abdullah Al Mamun ; Alma Islam 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)
15 Chatbot using TensorFlow for small Businesses Rupesh Singh ; Manmath Paste ; Nirmala Shinde ; Harshkumar Patel ; Nitin Mishra2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)
16 A Graph Based Chatbot for Cancer Patients R. V. Belfin ; A. J. Shobana ; MeghaManilal ; Ashly Ann Mathew ; BlessyBabu 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)
17 Doly: Bengali Chatbot for Bengali Education Md. Kowsher ; Farhana SharminTithi ; M Ashraful Alam ; Mohammad Nurul Huda ; Mir Md Moheuddin ; Md. Golam Rosul 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)
18 Chatbot and bullyfree Chat V. Selvi ; S. Saranya ; K. Chidida ; R. Abarna 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) Year: 2019 | Conference Paper | Publisher: IEEE
19 Utterance Censorship of Online Reinforcement Learning ChatbotYixuan Chai ; Guohua Liu 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI) Year: 2018 | Conference Paper | Publisher: IEEE
20 Query Expansion Evaluation for Chatbot Application Laila Hidayatin ; Faisal Rahutomo 2018 International Conference on Applied Information Technology and Innovation (ICAITI) Year: 2018 | Conference Paper | Publisher: IEEE.
21 Model of Multi-turn Dialogue in Emotional Chatbot Chien Hao Kao ; Chih-Chieh Chen ; Yu-Tza Tsai 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI) Year: 2019 | Conference Paper | Publisher: IEEE.
22 Chatbot for university related FAQs Bhavika R. Ranoliya Nidhi Raghuwanshi ; Sanjay Singh 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
23 Developing Smart Workspace Based IOT with Artificial Intelligence Using Telegram ChatbotMuhamad Muslih ; Somantri ; DediSupardi ; ElpidMultipi ; Yusup Maulana Nyaman ; Aditya Rismawan ; Gunawansyah 2018 International Conference on Computing, Engineering, and Design (ICCED) Year: 2018 | Conference Paper | Publisher: IEEE
24 A Cognitive Model for Emotion Awareness in Industrial ChatbotsAchiniAdikari ; Daswin De Silva ; DammindaAlahakoon ; Xinghuo Yu 2019 IEEE 17th International Conference on Industrial Informatics (INDIN) Year: 2019 | Volume: 1 | Conference Paper | Publisher: IEEE.
25 Chatbot for Disease Prediction and Treatment Recommendation using Machine LearningRohit Binu Mathew ; Sandra Varghese ; Sera Elsa Joy ; Swanthana Susan Alex 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)
26 Chatbot Implementation for ICD-10 Recommendation System NopponSiangchin ; TaweesakSamanchuen 2019 International Conference on Engineering, Science, and Industrial Applications (ICESI) Year: 2019 | Conference Paper | Publisher: IEEE

AUTHORS PROFILE

Saritha A K received the B.Tech degree in Computer Science & Engineering from Calicut University in 2009, the M.Tech degree in Computer science and Engineering from Calicut University, in 2012. She is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Gitam University, Bangalore. Her current research focuses on Cyber Security, IoT.