ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter169872020.pdf
https://doi.org/10.30534/ijeter/2020/169872020

Design of a Web Application for the Detection of Diabetes using Machine Learning

Alexi Delgado[1], Enrique Lee Huamaní[2], Victor Islachin-Minchola[3]
[1]Mining Engineering Section, Pontificia Universidad Católica del Perú, Lima, Perú
[2]Image Processing Research Laboratory, Universidad de Ciencias y Humanidades Lima, Perú
[3]Systems Engineering Program, Universidad de Ciencias y Humanidades, Lima, Perú

ABSTRACT

Diabetes is a disease that has always had an impact on humanity because of its terrible effects. Machine learning is the method used in this work, which allows us, through its classifiers, to predict upcoming events, in this case, probability of disease. The case study is Lima, Peru, where most cases of diabetes are seen, also using a chat Bot, to represent the doctor in the consultation. The result of this research was the design of this online doctor, which is used for the best care of patients and predicts diabetes, thus helping people who need a consultation. This research was designed for its next development, in order to serve as a starting point for other research with different diseases.

Key words - Decision Tree; Chat Bot; Diabetes; Machine Learning

1. INTRODUCTION

Today, diabetes has become one of the most important and recognized diseases in the world, not only because of its terrible effects on the human body, but also because of its difficult previous diagnosis, according to medical data it is estimated that this disease will increase from 376 billion to 490 billion in 2030 [1], in another aspect, diabetes presents some symptoms, such as continuous urination, increased thirst, expanded craving and weight reduction [2]. Diabetes can be divided into several categories among the main ones are type 1, it is caused due to total insufficiency of insulin secretion and type 2 is caused due to the combination of immunity to insulin action and insulin deficiency [3].

In the methodology was used, for the prediction of diabetes, Machine Learning (ML), which is a subfield of artificial intelligence, automatic learning has advanced by the need to instruct the PC, how to automatically take a response [2], in the use of Machine Learning, there are different classifiers, in the area of prediction, various classifiers are used for example; the first is the network of bays, this methodology tries to calculate the conditional probability of class and then predict the most probable class [4], it is widely used in various investigations, another method is data mining, this method performs the process of extraction of hidden patterns, hitherto unknown from huge database or data storage [5], methodology that in the present years has become very recognized, another very interesting methodology is the method of joint learning, random forest (RF) and the extreme drive gradient (XGBoost), which come together for better accuracy and classification [6], and finally, the decision tree classifier, this classifier contains control statements, which are used to select the appropriate resolution [7].

In the application, we used data from a data set acquired from Kaggle, called Pima Indias, extracted from natives of Arizona - USA. In the US, the decision tree will be used as the basis for the design, but from a different point of view, a web page will be designed that will have a chatbot, so that it will be the intermediary between the patient and a virtual doctor who will consult you from the data you give it, all Chatbot have the ability to answer the questions of some fields and particular questions established in accordance with it [8], consequently served to collect the information necessary for the prediction.

The objective of this research work is to prevent cases of diabetes in Lima through the design of a web software with Machine Learning, which allows to obtain a percentage of probability of diabetes, through a ChatBotMedic.

The present research work is structured in the following way, in section II, we will make known the method to design the web software and the materials that will be used to design the algorithm to obtain the probability, in section III this methodology proposed in the previous section will be developed, in section IV the results obtained will be shown based on surveys of the use of the software and the efficiency of the algorithm, and finally in section V, the conclusions will be given and the results obtained from this research will be discussed.

2. METHODOLOGY

In this research Machine Learning was used as the basis for the methodology of all the work, from this subfield of artificial intelligence was used the decision tree classifier, then we will mention the tools that were used and we will also know the steps for the design of the online query through Chatbot.

2.1 Machine Learning

Automatic learning (ML) is a computational method for automatic learning of experience, and improves performance to make more accurate predictions [9].

2.2 Decision Tree Classifier

A decision tree consists of tests or attributes nodes linked to two or more sub-trees and leaves or decision nodes marked with a class meaning decision [10].

2.3 Python

Python provides an easy to use interface that allows us to write directly into an interactive interpreter our Python programs will be carried out by the program directly [8].

2.4 ChatBot

A fundamental form of communication between the program and the code [8], it can be developed in different programming languages and serves as an aid in interaction.

2.5 Steps for the creation of ChatBotMedic

A. Definition of the database

When we talk about definition, we are referring to how data is used and how it is classified, what values each column has and how each field is used for classification.

B. Definition of the database

When we talk about definition, we are referring to how data is used and how it is classified, what values each column has and how each field is used for classification.

C. Creation of the algorithm with the chosen method

When we talk about definition, we are referring to how data is used and how it is classified, what values each column has and how each field is used for classification.

D. Building the Chatbot

In this step, already having all the logical part, for the design a Chatbot will be used, so that it is the one that asks the pertinent questions and can use the algorithm and give the answer, for it we will use 5 steps for its creation and feedback of the software. The first step is the need or commonly known as the requirements or what I need for my application [11]. In this case we will use the user stories to be able to do it in the Figure 1 Writing User Stories, shows us how to do it.

Como (rol) Necesito (funcionalidad) Para (beneficio)[24]

Figure 1: Writing User Stories

In the imagination of a software, various programs are used for the layout or design so that in this part you model or design what you want to see in the final software. In the creation itself the use of programming tools and other frameworks is made for the creation of the software and to make tests of various types, to ensure the reliability of the software, as it will have to be very fast and easy to use. It will be agile; the tests will be taken and the software will be improved quickly according to what fails.

E. Validation of the system

Finally, tests will be carried out on the software when implemented, to see how it works with other people, and will be tested with data from people with and without diabetes. The design methodology presented above consists of 5 steps that can be seen graphically in Figure 2 and will be developed in the next chapter.

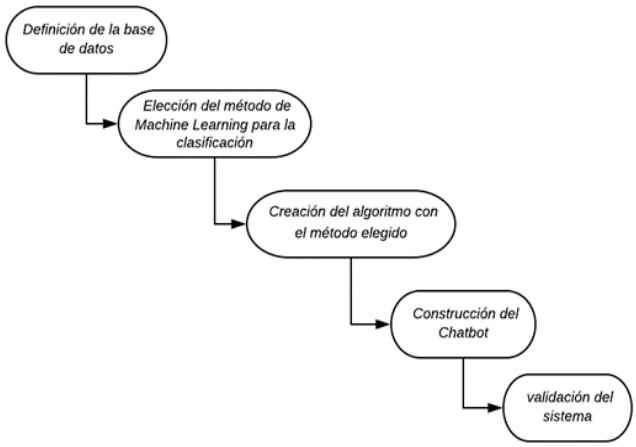

Figure 2: Steps in creating the software

3. APLICATION

As we can see in the methodology, we have talked about all the technologies that were used, but we will focus on the part of the software design, which is in point E that is developed below.

3.1 Definition of the database

There are several sources of data for our design, which gives us various fields that are useful, but not all are specific to our research, we have selected a data set that has a lot of order and gives us very important elements for the measurement of diabetes, shown in Table I, for example we have the number of pregnancies (a1), glucose (a2), blood pressure (a3), skin thickness (a4), insulin (a5), BMI (a6), diabetes pedigree function (a7), age (a8), result (a9).

3.2 Choice of the Machine Learning method for the classification

For the design of ChatBot we used the decision tree shown in Figure 3, as the classifier for our data, since it will give us, when implemented, an answer if there is probability of diabetes or not, then we will seek to improve the algorithm, so that it can give a probability along with the answer.

3.3 Creation of the algorithm with the chosen method

In the creation of the algorithm was used as an IDE, to Jupyter Notebook, as it was investigated that it is one of the most powerful IDE for information management and for Python.

Table 1: Diabetes Data, Source: Kaggle

(a1)	(a2)	(a3)	(a4)	(a5)	(a6)	(a7)	(a8)	(a9)
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0	54	1
4	110	92	0	0	37.6	0	30	0

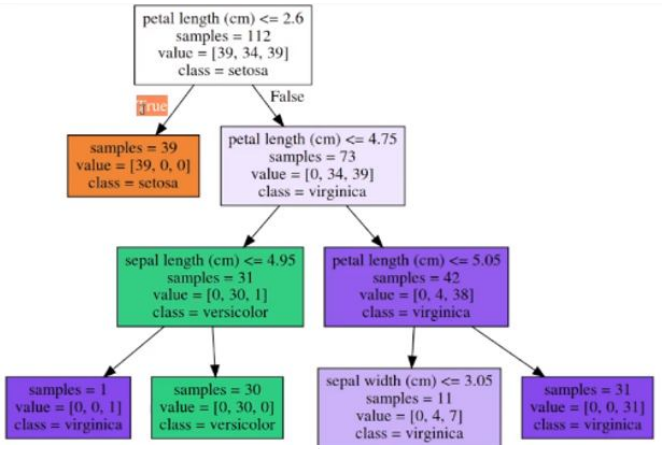

Figure 3: Example of decision tree algorithm

3.4 Construction of chatbot

In design, we use user stories as the basis for ChatBot.

- As a Patient I need to consult my disease to calculate the rate of diabetes.
- As a Patient I need to consult my disease to calculate the rate of diabetes.
- As ChatBotMedic I need to register questions to be able to ask the patient.
- As ChatBotMedic I need to register answers to be able to use them in automatic learning.

For the imagination part we created a small model shown in Figure 4, made in a rustic way made in Paint, without the need of Balsamiq as it will only be used as a model.

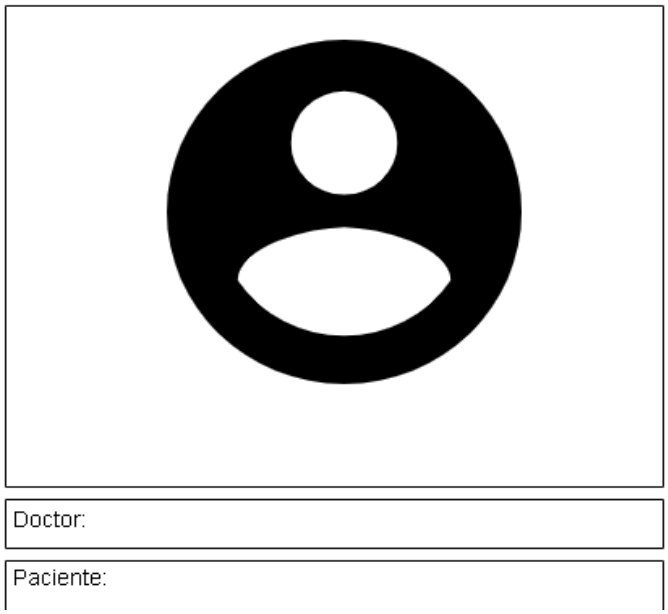

Figure 4: Prototype carried out

In the beginning of the creation, as it is only the design, the Python programming was thought using the Django framework for the speed at the time of making the Chatbot. In Figure 5, the following prototype was designed for the pair testing of ChatBotMedic. Finally, some users were shown to see the reaction and how the impact on the implementation of this software would be.

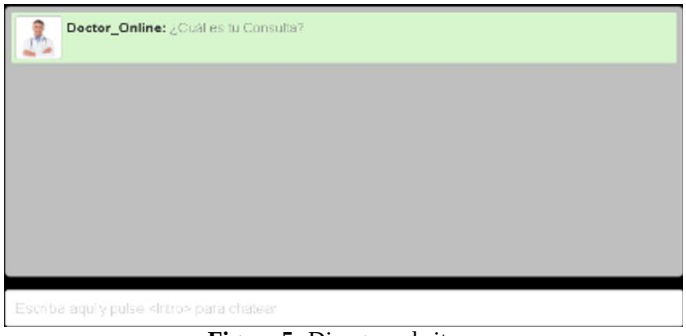

Figure 5: Django website

3.4 System validation

System design validated, but in the implementation will be validated by professionals such as doctors in diabetes and also with people who have this disease, will be tested with diagnostic people and people who do not have any disease.

4. RESULTS AND DISCUSSION

4.1 About the Case Study

In the case study, the design of the web system was achieved with a Chat Bot, especially to apply it to the detection of diabetes with Machine Learning. In the design of Figure 6, you can flow the future functionality that the Chat Bot will have.

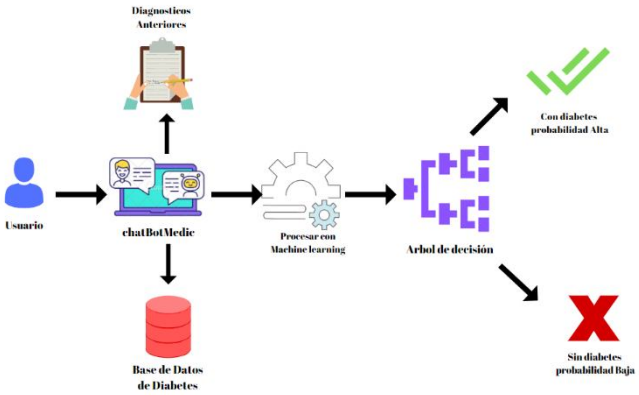

Figure 6: Software flow

As we can see, finally in the flow will have a ChatBotMedic with a graphical interface, which will serve as the mediator between the data, the user and the automatic learning code, to be able, through the answers you can provide, give a probabilistic percentage if you have diabetes or do not have it. There is no research clearly equal to this one, but compared with other research, was the design and implementation of a Chatbot with machine Learning, but with a commercial approach [12], since normally, people use chats so that customers are always connected to the company, compared with this research, it is desired that patients have great acceptance of this software, through its easy use, as well as customers usually have that reception with the Chatbots.

4.2 About the Methodology

In this investigation, remembering, the methodology of machine Learning, which will serve for the design of this project, below are mentioned the advantages and disadvantages.

A. Advantages

On the methodology, the advantages it has the machine Learning are that allows to predict situations that may arise in the future, and these predictions in turn allow a great benefit in time, business and human life [13].

B. Disadvantages

In my opinion, it takes a lot information so that the predictions can work, in addition that it is necessary to clean the data correctly, since it has a percentage of error by those dirty data.

In this investigation, using machine Learning, the decision tree classifier was used, in other investigations the support vector machine is used, a technique that has a good percentage of accuracy, in this investigation an accuracy of 84.1% was achieved [14][15][16], so we can say that it is also a good classifier, but we expect the decision tree to have a higher percentage of accuracy than shown by this investigation.

5. CONCLUSIONS

The design of this Chabot, will have a very positive impact on the Lima community, since it will produce improvements to the way in which normally a medical consultation is made, and by the rapidity in the results, from a point of view, also this project would help the ministry of health and other national centers, which have a large number of patients with this disease.

The machine learning method has proved to be very useful, in predictions, and is also widely used in health. Another positive point of this method is that it allows you to see the success rate of your algorithm.

Finally, in a future research, all the development will be done, and to add, that this research could be used for others, using another disease.

REFERENCES

[1] F. Faruque, “Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus,” 2019 Int. Conf. Electr. Comput. Commun. Eng., pp. 1–4, 2019. https://doi.org/10.1109/ECACE.2019.8679365

[2] D. Dutta, D. Paul, and P. Ghosh, “Analysing Feature Importances for Diabetes Prediction using Machine Learning,” 2018 IEEE 9th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2018, pp. 924–928, 2019.

[3] K. Sumangali, B. S. R. Geetika, and H. Ambarkar, “A classifier based approach for early detection of diabetes mellitus,” 2016 Int. Conf. Control Instrum. Commun. Comput. Technol. ICCICCT 2016, pp. 389–392, 2017.

[4] Y. Guo, G. Bai, and Y. Hu, “Using Bayes Network for prediction of type-2 diabetes,” 2012 Int. Conf. Internet Technol. Secur. Trans. ICITST 2012, pp. 471–476, 2012.

[5] F. G. Woldemichael and S. Menaria, “Prediction of Diabetes Using Data Mining Techniques,” Proc. 2nd Int. Conf. Trends Electron. Informatics, ICOEI 2018, no. Icoei, pp. 414–418, 2018.

[6] Z. Xu and Z. Wang, “A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier,” 2019 Elev. Int. Conf. Adv. Comput. Intell., pp. 278–283, 2019.

[7] D. Vigneswari, N. K. Kumar, V. Ganesh Raj, A. Gugan, and S. R. Vikash, “Machine Learning Tree Classifiers in Predicting Diabetes Mellitus,” 2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019, pp. 84–87, 2019. https://doi.org/10.1109/ICACCS.2019.8728388

[8] B. Kohli, T. Choudhury, S. Sharma, and P. Kumar, “A Platform for Human-Chatbot Interaction Using Python,” 2018 Second Int. Conf. Green Comput. Internet Things, pp. 439–444, 2019.

[9] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," Appl. Comput. Informatics, no. xxxx, pp. 0–5, 2019.
[10] E. Papageorgiou, C. Stylios, and P. Groumpos, "A combined fuzzy cognitive map and decision trees model for medical decision making," Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc., pp. 6117–6120, 2006.
[11] M. Trigás Gallego, "Metodología Scrum," Gest. Proy. informáticos, p. 56, 2012.
[12] P. Kumar, M. Sharma, S. Rawat, and T. Choudhury, "Designing and Developing a Chatbot Using Machine Learning," 2018 Int. Conf. Syst. Model. Adv. Res. Trends, pp. 87–91, 2019.
[13] H. I. Bülbül and Ö. Ünsal, "Comparison of classification techniques used in machine learning as applied on vocational guidance data," Proc. - 10th Int. Conf. Mach. Learn. Appl. ICMLA 2011, vol. 2, pp. 298–301, 2011. https://doi.org/10.1109/ICMLA.2011.49
[14] H. Abbas, L. Alic, M. Rios, M. Abdul-Ghani, and K. Qaraqe, "Predicting Diabetes in Healthy Population through Machine Learning," pp. 567–570, 2019.
[15] A. Delgado, P. Montellanos, and J. Llave, "Air quality level assessment in Lima city using the grey clustering method," IEEE ICA-ACCA 2018 - IEEE International Conference on Automation/23rd Congress of the Chilean Association of Automatic Control: Towards an Industry 4.0 – Proceedings, 8609699, 2019.
[16] A. Delgado, and I. Romero, "Applying the Grey Systems Theory to Assess Social Impact from an Energy Project," Proceedings of the 2018 IEEE 25th International Conference on Electronics, Electrical Engineering and Computing, INTERCON 2018, 8526372, 2018. https://doi.org/10.1109/INTERCON.2018.8526372