Mini Review

Explainable AI in Diabetes Prediction System

Dr. Jagreet Kaur[1], Suryakant[2] and Kuldeep Kaur[3]*

[1]*Chief AI Officer at (XenonStack Private Limited), Founder and CEO at (Xenon DigiLabs Private Limited), AI and Analytics Department, Xenonstack, Punjab, India*
***E-mail:** jagreet@akira.ai*
[2]*ModelOps Specialist, AI and Analytics Department, Xenonstack, Punjab, India*
[3]*AI Ethics Researcher, AI and Analytics Department, Xenonstack, Punjab, India*

***Corresponding Author:** Kuldeep Kaur, AI Ethics Researcher, AI and Analytics Department, Xenonstack, Punjab, India.*

Abstract

Use of AI in healthcare improves the industry services. Discovering patterns from data using ML improves the decision making process. It allows the industry specialist to make data-driven and fact-based decisions. The use of ML models in Healthcare is continuously increasing but it proliferates the concerns of stakeholders due to complexity and black box functioning of ML models.

Therefore Explainable AI approaches come into existence to make the ML model transparent and trustworthy. In this document a case study is represented in which the ML model is used to detect diabetes and for transparency Explainable AI approaches are defined to understand the AI system based on the concerns and queries [1] that would be raised by stakeholders. There are several approaches, libraries and packages that can be used to implement Explainable AI such as LIME [2], SHAP [3] etc. It allows the industry practitioner to use the right tools and approaches for making their AI system trustworthy and transparent.

Keywords: Diabetes; Healthcare; AI System

Challenge of AI system

The main challenge in an AI system is a customer's trust. Opaque AI systems provide the output of the system without giving any reason or Explanation. Therefore, It becomes challenging for the customer to trust the machine that does not give any explanation, especially in the healthcare system. Various questions come to the customer's mind that the opaque system is not able to answer, as figure 1 shows. Due to this incompetence of opaque AI systems, they are not adopted by the patients and the medical practitioners.

Solution by Akira AI

To overcome the challenge, Akira AI explains their opaque AI systems. It answers the questions that arise in the customer's mind while using the AI system. It makes an AI system more reliable and productive by providing trust, transparency, and fidelity. To answer all these questions, Akira AI uses Explainable AI.
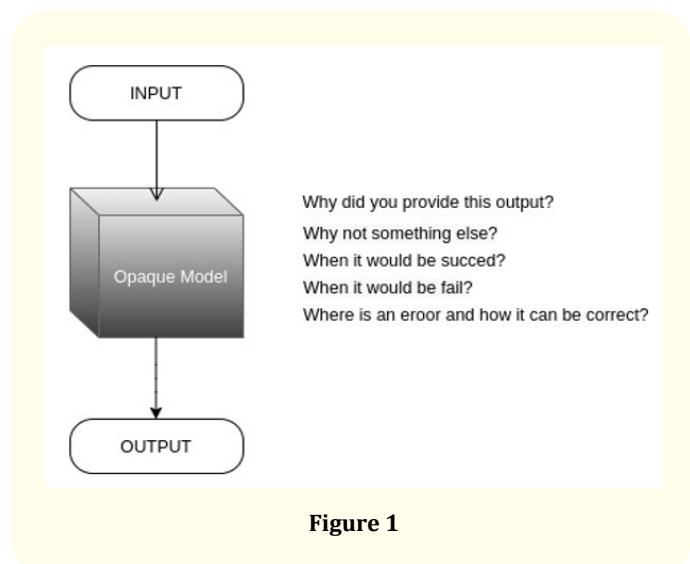

Figure 1

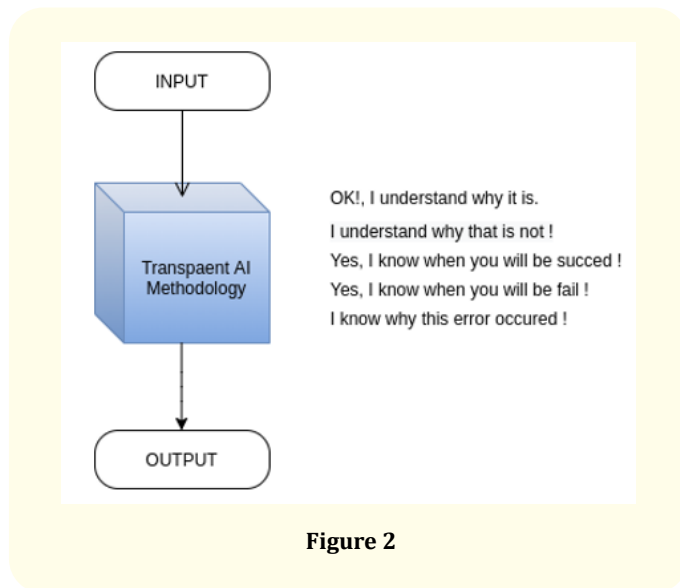Figure 2 depicts that the customer can answer a question that helps them understand the model and its working.

Figure 2

Explainable AI in healthcare

Why, What, How, Where and When of Explainable AI in Healthcare.

Why do we need explainable AI in Healthcare?

Healthcare is an industry where it is necessary to always take a precise decision, whether it is a treatment, test, or discharge. Any neglect can have a high cost for the patients as well as for the medical practitioner. It becomes challenging for the patient to trust when they get to know that this decision is taken by the machine that is not explaining how it reaches a particular conclusion.

So the use of the Explainable AI is mandatory in predicting disease that will help gain the confidence of an AI system result.

Explainable AI helps to get the fair and correct output without errors.

What type of explanation do we need in healthcare?

Explainable AI is not just about giving justification for the model's decision. It is worth more than that. Explanation of the output is not enough to solve all the queries of the customer. Akira AI explains each element of the solution process. These are

- **Data:** It explains the data used for the prediction, their correlation, and EDA to understand the hidden data patterns. It tells how the data is to be used for the AI system [4]

- **Algorithm:** A complete transparency of the algorithm used by the system is given with the reason why the system chooses it and how it can be beneficial for the prediction [4]

- **Model:** Akira AI gives a detailed explanation of model, its performance and working in a user-friendly manner [4]

- **Output:** Akira AI gives a complete justification for the system's output with the reason. It also provides the factors that contribute to influence the result of the system [4]
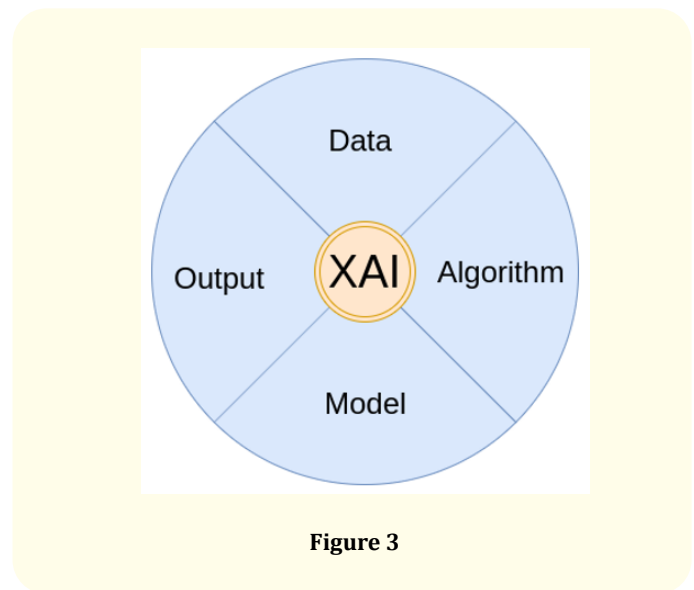

Figure 3

How to implement explainable AI in healthcare?

There is a step by step approach to implement the Explainable AI. From selecting data to the deploying system, Akira AI has methodologies and framework that can be implemented for providing the Explanation, such as

- Selection of the right data and preprocess it for model [4]
- Precisely select the suitable algorithm for the system [4]
- Use various frameworks to make the model working transparent [4]
- Justify the model output and factors on which it depends [4]
- To use a clear and clean deployment process [4].

Where to implement explainable AI in healthcare

It is not required to use the Explainable AI everywhere in Healthcare. There are some cases where it is necessary to maintain

privacy; it can be regarding the data and model working. Therefore we have to use Explainable AI precisely only at required places.

We can use it when we need to give assurance to patients and doctors. Such as to prescribe medicines, to diagnose disease, etc.

Different approaches can be used for different types of Explanations. We can implement Explainable AI in processes that need performance, trust, and confidence.

When of explainable AI in healthcare

There is not a time limit for providing Explanation to their system. It cannot be said that we require Explainable AI now, and it will be of no use in the future.

It is a requirement of the past, present, and future. But as we have already discussed, it is not compulsory to use it everywhere. It can be used only for some specific use cases where it is mandatory to have transparency.

Case study-diabetes prediction system

A system is used to predict whether or not a patient has diabetes based on some of its health-related details such as BMI, blood pressure, insulin, etc. This system is only for females as the dataset that is used to make this system exclusively belongs to the females.

Here, it is required to have an accuracy of the model to predict the correct result, as it is about someone's health. A single wrong result can have an adverse effect. Therefore, a good accuracy "Random Forest Classifier" model is selected, which provides good accuracy to the system and generates the correct results.

Transparency is equally essential as accuracy so that doctors, as well as the patient, can check whether the output that the system generated is accurate or not. But as we know Random Forest classifiers belong to the opaque model family and not providing much transparency makes it difficult to understand the complex working of the algorithm.

Therefore to balance, both the performance and interpretability of Akira AI, use Explainable AI here.

End customer questions

End customer always has some question in their mind regarding the AI system and its working. Akira AI answers all those ques-

tions using various Explainable AI methodologies. Some of these question that comes in the customer's mind by looking at the Diabetes Prediction AI system

- How data is contributing in making a decision? [1]
- Which feature influences the result more? [1]
- How changing the value of a feature changes the system output? [1]
- How could the system say that I can have diabetes in the future? [1]
- What is the amount of the Glucose that I need to maintain? [1]
- How much sugar and insulin should I have to decrease to keep my sugar level? [1]
- What is the general rule that the system used to generate an output? [1]

Which feature influences the result more?

It can build the trust of doctors and patients on the system's results.

The figure 4 depicts the importance of the features in predicting the output. Features are sorted in decreasing importance from top to bottom to generate the output.
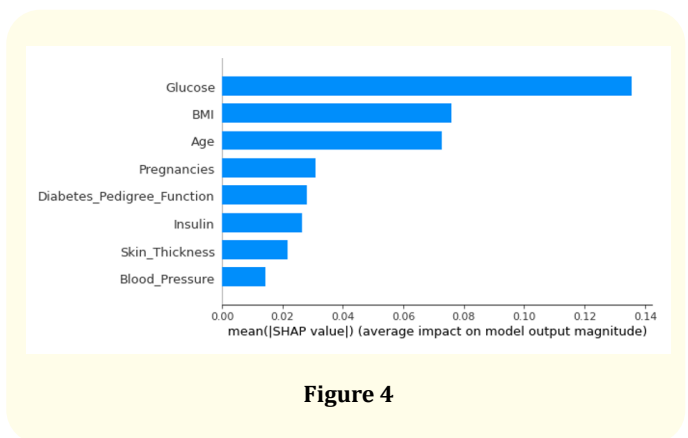

Figure 4

As it is shown, the Glucose value of a person influences the result more while predicting whether a person can have diabetes or not.

How data is contributing to making a decision?

This is the next version of the previous graph. It also shows the

same things with some more information about the value of the feature.

- **Feature importance:** Variables are ranked in order of their importance such as from top to bottom importance reduces.

- **Impact:** The horizontal location shows the effect of that value associated with a higher or lower prediction.

- **Value:** Color shows the value of variable whether that variable is high or low for an observation. The red color represents the high value while blue for less value. The variation in color of the dot shows the value of the feature

- **Correlation:** A high level of "Glucose" content has a high impact on having diabetes. The "high" comes from the red color, and the effect is shown on the X-axis.
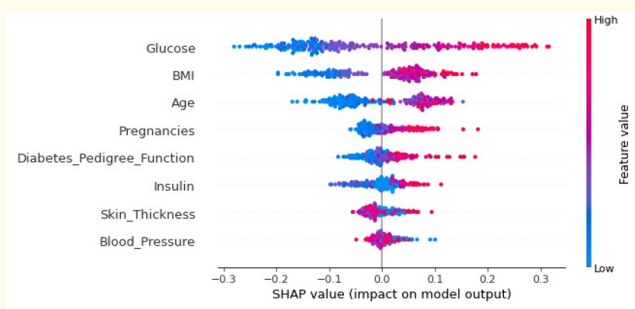

Figure 5

How does a change in the value of the Glucose change the system output?

- After getting the answer to the first question, the customer can ask how the change in the value of the Glucose changes the system output when other parameters are not changing?

- So, let's discuss the Partial Dependence Plot (PDP) [5]

- PDP shows the relation between the target response and feature. Other features are marginalized. This graph shows how a change in the value of Glucose changes the predicted output. It depicts that the increase in Glucose's value increases the probability of having diabetes.

ICE (Individual Conditional Expectation) [6] is a more detailed view of PDP. ICE is used to inspect the model's behavior for a specific instance, where everything except Glucose is held constant, fixed to their observed values. At the same time, Glucose is free to attain different values.
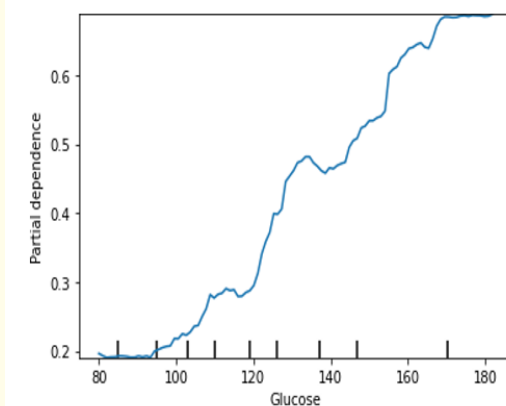

Figure 6

In PDP, other features are averaged out, but it took each case individually and the plot graph by only changing salary and remaining others constant. So this is a broad view of PDP.

In figure 7, we just selected some of the observations. Users are free to choose any number of observations that they want to explore.
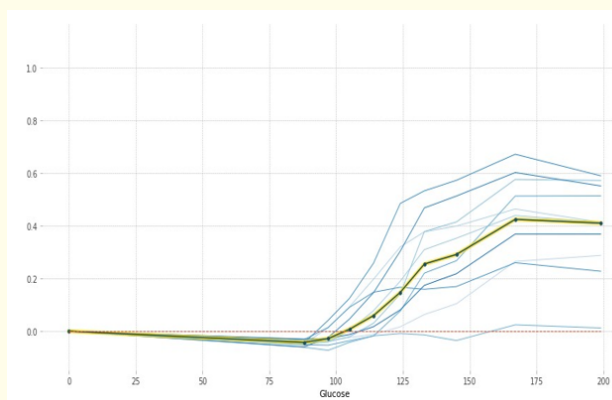

Figure 7

Why did the system predict I can have diabetes in the future?

Using SHAP [3], the system plots and justify its result is why it chooses individual output. The SHAP value shows the impact of feature evidence on the model's output.

The waterfall plot depicts how each feature's SHAP values move the model output from our prior expectation under the background data distribution to the final model prediction. It gives evidence of all the features.

It displays explanations for individual predictions, so they expect a single row of an Explanation object as input. Each row shows how the positive (red) or negative (blue) contribution of each feature moves the value from the expected model output over the background dataset to the model output for this prediction.
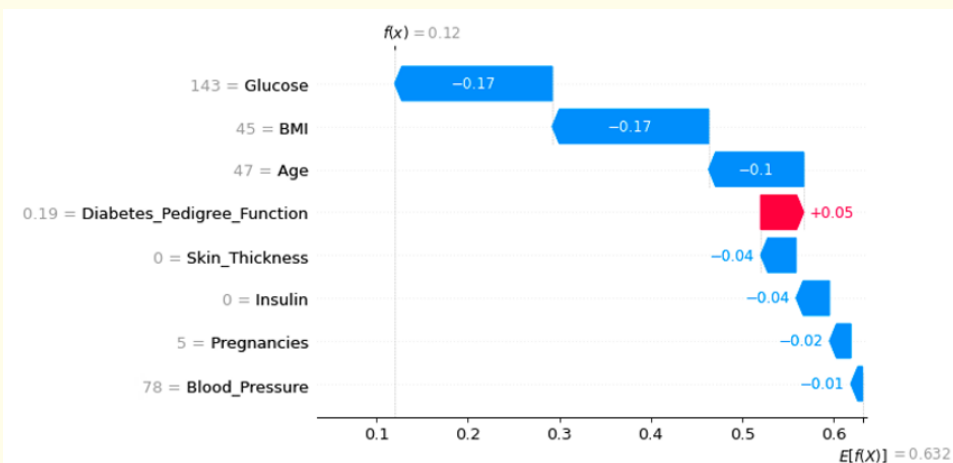

Figure 8

What is the average value of the Glucose I need to maintain?

• The glucose value of the person plays a significant part in predicting whether a person can have diabetes or not.

• Anchors can answer what Glucose's value that an individual needs to maintain to not to have diabetes is.

• So this figure depicts that if that person maintains the Glucose value less than 100, it cannot have diabetes.
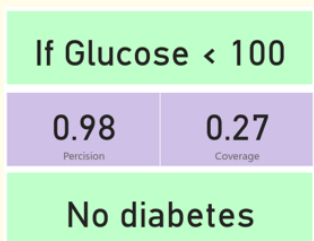

Figure 9

What is the general rule the system used to generate an output?

• This is an essential question. A satisfying answer to that question can clear many doubts about the customers and bring their confidence in the system. If the system can provide the general rule that the system obeys to give the output, it can clear all the doubts regarding the algorithm's working.

• We can use various methodologies to answer the question according to the algorithm we used to predict the output as we are using the Random Forest Classifiers Algorithm. Hence, we have some methods such as inTRess, defragTrees, etc. that we can use to extract the Random Forest Algorithm rule. Here we used defragTrees method that generates rules that the system used to predict an output as shown in Figure 1.10.

• It also provides the performance attributes from which the user can analyze and track the algorithm's performance and rule to predict an output. It tracks the model's behaviour on

a global scale. Thus inspect it to find out whether the model has picked up any undesired functioning.pect it to find out whether the model has picked up any undesired functioning.

Figure 10

Figure 10 describes two rules on which system is working. It provides the value of the parameters that helps to reach the system at the output. As given in Rule 1 if the value of the Glucose <= 100 and BMI <= 30 and Age <= 30 and Pregnancies <= 4 and Skin thickness <= 35 and Insulin <= 35 then the system will say that a lady doesn't have diabetes. Whereas opposite of these values as given in the Rule 2 if the values of the parameters are Glucose > 100 and BMI > 30 and Age > 30 and Pregnancies > 4 and Skin thickness > 35 and Insulin > 35 then it will say that the lady can have diabetes.

It also provides the Test Error, Test coverage and the Overlap values that helps to understand the performance of the system.

Conclusion

The contribution of the Explainable AI in the Diabetes Prediction system makes it easy for the end-user to understand the AI systems and their complex working. It provides a human-centered interface to the user. Explain ability is a key to producing a transparent, proficient, and accurate AI system that can help the healthcare practitioner, patients, and researcher understand and use the system.
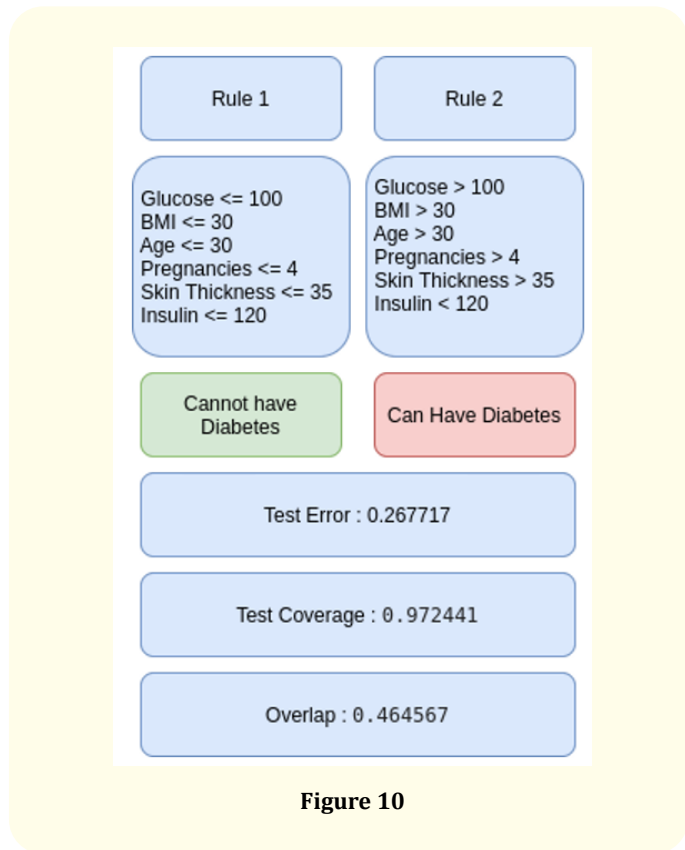
Bibliography

1. Vaishak Belle., *et al*. "Principles and Practice of Explainable Machine Learning". arXiv:2009.11698v1. 2020-09-18.

2. Scott Lundberg., *et al*. "LIME".

3. Marco Tulio Correia Ribeiro., *et al*. "SHAP".

4. Ankur Teredesai., *et al*. "Explainable Models for Healthcare AI". KenSci., Inc. (2018).

5. Christoph Molnar. "Partial Dependence Plot". 2021-08-22.

6. Christoph Molnar. "Individual Conditional Expectation". 2021-08-22.

Volume 5 Issue 10 October 2021