Project Design Phase-I

DATE	26/09/2023
PROJECT NAME	AI Based Diabetes Prediction System

PROBLEM STATEMENT:

The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyze medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their health.

PROPOSED SOLUTION

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.NO	PARAMETER	DESCRIPTION
1.	Problem Statement (Problem to be solved)	Many individuals are at risk of developing diabetes, but early detection is often challenging. A solution is needed to accurately predict diabetes risk factors and enable early intervention to prevent the onset of the disease.
2.	Idea / Solution description	The AI-based Diabetes Prediction System utilizes machine learning algorithms and patient data to predict an individual's risk of developing diabetes.
3.	Novelty / Uniqueness	This system stands out by leveraging cutting-edge machine learning techniques to provide accurate and personalized diabetes risk predictions. It takes into account a wide range of factors, including genetic markers, which adds a unique dimension to early prediction.
4.	Social Impact / Customer Satisfaction	By identifying individuals at risk of diabetes at an early stage, this system can significantly contribute to public health. It empowers individuals to make informed lifestyle choices and enables healthcare providers to offer targeted preventive interventions. Ultimately, it improves overall customer satisfaction by promoting health and well-being.
5.	Business Model (Revenue Model)	The business model revolves around subscription-based access to the Diabetes Prediction System. Users can access basic risk assessments for free, while premium features, such as more detailed reports and personalized recommendations, can be unlocked through a subscription fee.
6.	Scalability of the Solution	The AI-based Diabetes Prediction System is designed to be highly scalable. It can handle a large volume of patient data and continuously improve its predictions as more data becomes available. This scalability ensures that it can serve a growing user base and adapt to evolving healthcare needs.

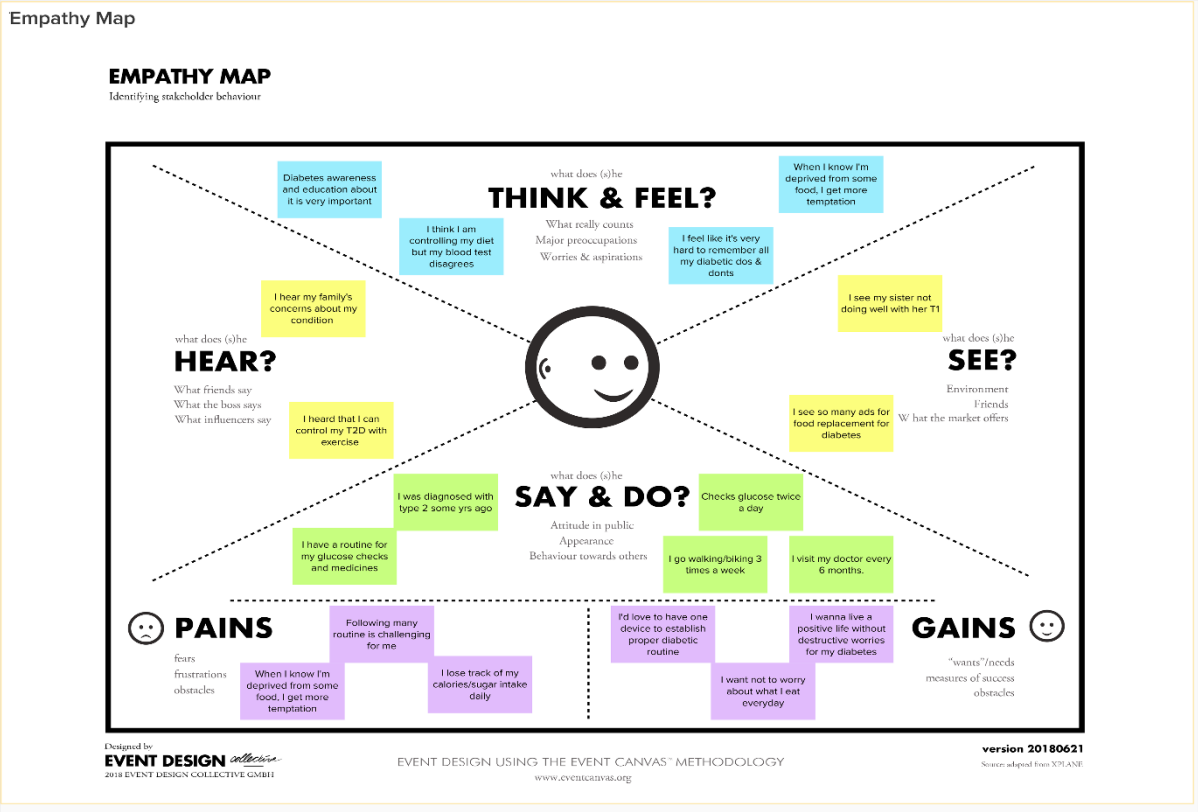
Ideation Phase

Empathize & Discover

Empathy Map Canvas:

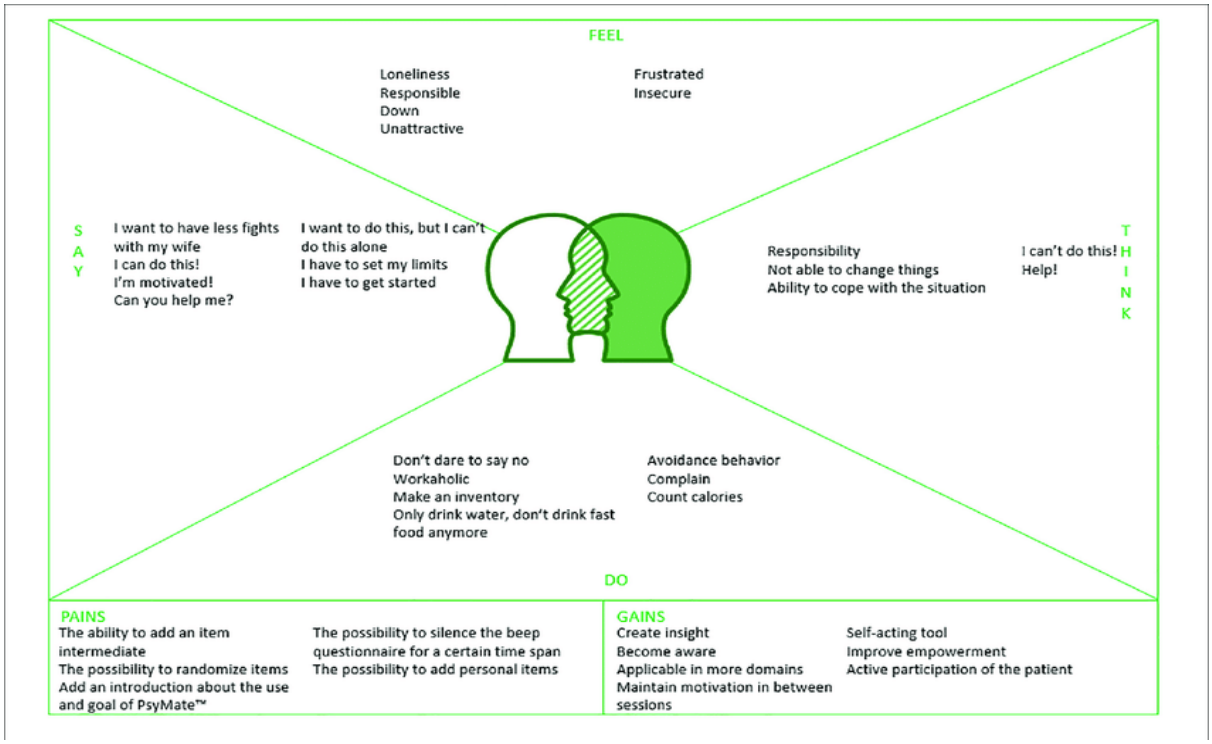
In creating an empathy map canvas for an AI-based diabetes prediction system, we focus on understanding the typical user, their expressed needs, thoughts, emotions, and actions. For instance, users might vocalize their struggles with blood sugar control, harbor concerns about long-term complications, experience frustration from constant monitoring, and engage in behaviors like blood sugar tracking and medication adherence. This canvas provides a holistic view of the user's perspective, helping guide the development of a more user-centric and effective diabetes prediction system.

Example



1. **User Persona:** Start by defining a representative user persona, such as a middle-aged individual with type 2 diabetes.

2. **Label "Says":** Create a section labeled "Says" where you note down direct statements or quotes from the user regarding their diabetes management. For example, "I struggle with managing my blood sugar levels."
3. **Label "Thinks":** In this section, jot down the thoughts, concerns, or worries that the user may have about their diabetes. For instance, "I'm concerned about the long-term effects of diabetes."
4. **Label "Feels":** Describe the emotions and feelings that the user experiences in relation to their diabetes. For example, "I feel overwhelmed by the constant monitoring and lifestyle changes."
5. **Label "Does":** Outline the actions and behaviors the user engages in concerning diabetes management. This could include activities like checking blood sugar, taking medications, and making dietary choices.
6. **Environment:** Consider the user's physical environment, including where they live and work, as it can influence their diabetes management.
7. **Influences:** Identify the factors and people that influence the user's decisions and behaviors, such as healthcare professionals, family, or friends.
8. **Pain Points:** List the challenges and difficulties the user faces in managing their diabetes, such as the complexity of tracking data or the fear of complications.
9. **Gains:** Note the user's goals, aspirations, and desired outcomes related to their diabetes, such as better control and improved quality of life.
10. **Takeaways:** Summarize the key insights gained from the empathy map that can inform the design and development of the AI-based diabetes prediction system, ensuring it addresses the user's needs, emotions, and actions effectively.



SOLUTION ARCHITECTURE

Solution Architecture:

Solution architecture is a structured approach to designing complex systems or projects, outlining the components, relationships, and processes to achieve specific goals or solve problems efficiently. It provides a high-level blueprint for project development and implementation.

- Gather and preprocess data from various healthcare sources, ensuring data quality.
- Train AI models for diabetes prediction and evaluate their performance.
- Deploy models in a secure, scalable environment with a user-friendly interface.
- Continuously monitor and update the system while complying with regulations.
- Collaborate with healthcare professionals and gather user feedback for improvements.

Example - Solution Architecture Diagram:

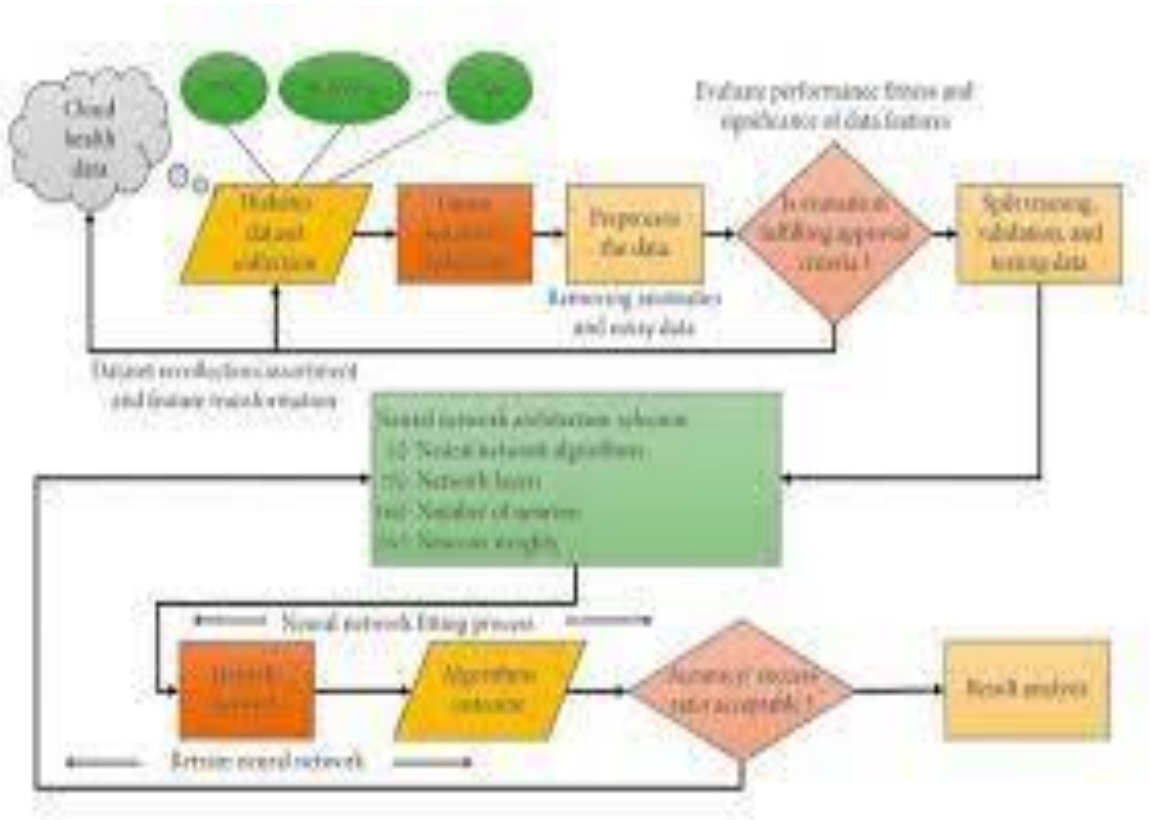


Figure: AI diabetes based prediction system

Reference: https://www.xenonstack.com/blog/artificial-intelligence-diabetes-detection

Literature Review

SI.NO	YEAR	PAPER	TITLE	AUTHOR	OUTCOME
1.	2017	IEEE	Application of Data Mining Methods in Diabetes Prediction	Messan Komi, Jun Li Yongxin Zhai, Xianguo Zhang.	The paper discusses the application of data mining methods for predicting diabetes risk. It explores the use of five different data mining techniques, including GMM, SVM, Logistic regression, ELM, and ANN, for early diabetes prediction. The paper concludes that ANN (Artificial Neural Network) provides the highest accuracy among these techniques.
2.	2019	IEEE	Prediction of Diabetes Patient Stage Using Ontology Based Machine Learning System	V.Swathi Lakshmi, V.Nithya, K.Sripriya, C. Preethi, K. Logeshwari	1.The paper proposes a method for identifying the risk level of diabetes in patients using ontology and machine learning techniques, particularly the Naïve Bayes algorithm. The proposed approach aims to improve the prediction of diabetes risk compared to existing methods. The authors discuss the use of ontology for holding disease-related data and the integration of semantic languages for identifying patients with diabetes. 2.The future work focus on applying and the some other technique to improve the better performance of these methods and efficient in time manner and also improve naïve bayes machine technique by applying some other methods with it.
3	2019	IEEE	Random Forest Algorithm for the Prediction of Diabetes	K.VijiyaKumar, B.Lavanya , I.Nirmala , S.Sofia Caroline.	The paper discusses the application of the Random Forest algorithm in predicting diabetes. Diabetes is a chronic disease with serious health implications, and early detection is crucial for effective management. The paper highlights that traditional diagnostic methods involve visiting a diagnostic center and consulting a doctor, which can be time-consuming and costly. The authors propose the use of machine learning techniques, specifically the Random Forest algorithm, to predict diabetes with high accuracy.

4	2020	IJERT	Diabetes Prediction using Machine Learning Techniques	Mitushi Soni, Dr. Sunita Varma	Random Forest-The accuracy it gives is grater then compared to other models .Random Forest Improve Performance of Decision Tree by reducing variance. It operates by constructing a multitude of decision trees at training time and outputs the class that is the mode of the classes or classification or mean prediction (regression) of the individual trees. In this project ,they have used Random Forest Algorithm to obtain the better accuracy. And 77% classification accuracy has been achieved.
5	2019	ICRTAC	Diabetes Prediction using Machine Learning Algorithms	Aishwarya Mujumdara , Dr. Vaidehi V	1)LOGISTIC REGRESSION :Gives 96% accuracy 2)PIPELINE: Application of pipeline gave AdaBoost classifier as best model with accuracy of 98.8%. To predict how non-diabetic people can have diabetes in next five years.
6	2023	IJSR	Diabetes Prediction Using Machine Learning	Author - KM Jyoti Rani	High Accuracy: One of the primary outcomes of the study is the achieved accuracy of 99% using the Decision Tree algorithm. This is a significant finding and suggests that the designed system has a high level of accuracy in predicting diabetes based on the dataset used. Future Directions: The study outlines future directions and opportunities for further research and improvement. These include automating diabetes analysis and incorporating additional machine learning algorithms to enhance the system's capabilities.
7	2020	IEEE	Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers	MD. KAMRUL HASAN ,MD. ASHRAFUL ALAM ,DOLA DAS, EKLAS HOSSAIN , (Senior Member, IEEE), AND MAHMUDUL HASAN .	The passage discusses the successful application of an ensemble model for diabetes prediction using the PID dataset. It underscores the crucial role of preprocessing techniques, particularly outlier rejection and missing value imputation, in improving dataset quality. These preprocessing steps enhance attribute distribution characteristics, such as kurtosis and skewness.

8	2023	IEEE	The Long-Term Effects of Physical Activity on Blood Glucose Regulation: A Model to Unravel Diabetes Progression	Pierluigi Francesco De Paola , Alessia Paglialonga , Pasquale Palumbo , Member, IEEE, Karim Keshavjee , Fabrizio Dabbene , Senior Member, IEEE, and Alessandro Borri .	Novelty of the Model: The proposed model is the first known attempt to establish a direct link between long-term benefits of physical activity and a specific protein, IL-6, in the context of diabetes progression, introducing a new factor in understanding this process. Future Directions: The research plans to implement model-based control techniques based on this novel model and validate it using data, such as the Diabetes Prevention Program data. Additionally, the study aims to explore the relationship between IL-6 and insulin resistance and characterize the benefits of physical activity under various scenarios, including different levels of insulin sensitivity decay and variations in exercise parameters.
9	2015	Crossmark	Early Detection and Treatment of Type 2 Diabetes Reduce Cardiovascular Morbidity and Mortality: A Simulation of the Results of the Anglo-Danish-Dutch Study of Intensive Treatment in People With Screen-Detected Diabetes in Primary Care	William H. Herman,1 Wen Ye,2 Simon J. Griffin,3 Rebecca K. Simmons,3 Melanie J. Davies,4 Kamlesh Khunti,4 Guy E.H.M. Rutten,5 Annelli Sandbaek,6 Torsten Lauritzen,6 Knut Borch-Johnsen,7 Morton B. Brown,2 and Nicholas J. Wareham 3	Early Diagnosis and Treatment Benefit: The results suggest that there are significant benefits associated with early diagnosis and treatment of glycemia and cardiovascular risk factors in type 2 diabetes. These benefits include substantial absolute risk reduction (ARR) and relative risk reduction (RRR) at 5 years when compared to scenarios with delayed diagnosis and routine treatment. Timing Matters More than Treatment Intensity: The study's conclusions emphasize that the timing of initiation of glucose, blood pressure, and cholesterol treatment after diagnosis is more critical than the intensity of treatment. Screening for type 2 diabetes can help reduce the lead time between diabetes onset and clinical intervention, potentially improving outcomes.
10	2017	IEEE	Predictive Modeling for presumptive diagnosis of Type 2 Diabetes Mellitus based on symptomatic analysis	Ordoñez Barrios, Diego Alberto ,Vizcarra Infantes, Erick Raphael,Armas Aguirre, Jimmy Alexander.	High Diagnostic Precision: The model exhibits a remarkable level of precision in diagnosing diabetes, particularly with the Auto Classification algorithm, which achieves an accuracy of 91.7%. This indicates that the model is effective at identifying patients with diabetes based on their symptomatic characteristics. Potential for Improvement and Integration: The model has the potential for further enhancement by using larger datasets collected in medical settings or from other institutions. Additionally, it is designed with scalability in mind and can benefit from real-

					time data collection through technological devices or Smart technology. Implementing a feedback mechanism can also help determine the true precision of medical diagnoses, further improving its accuracy and reliability.
11.	2021	IEEE	Typical and Non-Typical Diabetes Disease Predictionusing Random Forest Algorithm	Md. Tanvir Islam, M. Raihan, Fahmida Farzan, Nasrin Aktar	The purpose of this study is to identify the Diabetes Mellitus type accurately using Random Forest algorithm which is an Ensemble Machine Learning technique and we obtained 98.24% accuracy for seed 2 and 97.94% for seed 1 and 3.
12.	2023	IEEE	Impact of Nutritional Factors in Blood Glucose Prediction in Type 1 Diabetes Through Machine Learning	Giovaani Annuzzi, Pasquale Arpaia	A study to determine the effect of nutritional factors (i.e., carbohydrates, proteins, lipids, fibers, and energy intake) in the short and middle term on Blood Glucose Levels (BGLs) prediction was conducted by Machine Learning (ML) methods. AML model able to predict the BGLs after 15, 30, 45, and 60 minutes from the meal leveraging on insulin doses, blood glucose, and nutritional factors in T1D patients on AP systems was implemented.
13.	2021	IEEE	Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction	Nicos Fazakis , Otilia Kocsis , Elias Dritsas , Sotiris Alexiou, Nikos Fakotakis	A comparative study is presented among the Finnish Diabetes Risk Score (FINDRISC) and Leicester risk score systems and several ML models, using inductive and transductive learning. The experiments were conducted using data extracted from the English Longitudinal Study of Ageing (ELSA) database.
14.	2017	IEEE	Diabetes Prediction Using Ensemble Perceptron Algorithm	Roxana Mirshahvalad, Nastaran Asadi Zanjani	In this study, we proposed a learning algorithm which ensemble boosting algorithm with perceptron algorithm to improve performance of perceptron algorithm in prediction of undiagnosed patients. Proposed method is tested on three different publicly available datasets and compared with performance of perceptron algorithm. The results show that proposed algorithm outperform perceptron algorithm on average AUC basis

15	2016	IEEE	A First Attempt to Develop a Diabetes Prediction Method Based on Different Global Datasets	Anjili Neji,Varun JaisWal	In the present study a method is developed using combined datasets using machine learning technique. This system is more reliable because of trained, tested and validated on combine dataset.

References:

1. https://ieeexplore.ieee.org/document/7984706
2. https://ieeexplore.ieee.org/document/8878831
3. https://ieeexplore.ieee.org/document/8878802
4. https://www.ijert.org/research/diabetes-prediction-using-machine-learning-techniques-IJERTV9IS090496.pdf
5. https://www.sciencedirect.com/science/article/pii/S1877050920300557
6. https://doi.org/10.32628/CSEIT206463
7. https://ieeexplore.ieee.org/document/9076634
8. https://ieeexplore.ieee.org/document/10168196
9. https://doi.org/10.2337/dc14-2459
10. https://ieeexplore.ieee.org/document/8079667
11. https://ieeexplore.ieee.org/document/9225430
12. https://ieeexplore.ieee.org/document/10043722
13. https://ieeexplore.ieee.org/document/9491154
14. https://ieeexplore.ieee.org/document/8319383
15. https://ieeexplore.ieee.org/document/7913152