|  |  |
| --- | --- |
| Diabetes Prediction Web-App | Spandan Bandhu (Team Leader)  Souvick Mazumdar  Omar Mahmood  Tarasha Ahuja  Adarsh Singh |

**Problem Statement:**

The increasing prevalence of diabetes has created a pressing need for accessible and accurate prediction tools that can empower individuals to take proactive steps towards managing their health. Current methods of diabetes prediction are often cumbersome and require medical expertise, leading to barriers in early detection and intervention. As a result, there is a clear need for a user-friendly website that can accurately predict the likelihood of diabetes based on individual user data. This platform aims to bridge the gap between medical insights and everyday users, offering a convenient and reliable solution to enable early detection and informed decision-making regarding diabetes management.

**Business Need Assessment:**

The business landscape surrounding healthcare and wellness is rapidly evolving, with a growing demand for personalized solutions that cater to individual health concerns. This project addresses the need for a digital platform that leverages predictive analytics to offer users insights into their potential risk of diabetes. By providing an easy-to-use interface, backed by robust predictive models, this website can tap into the burgeoning market of health-conscious individuals seeking proactive health management tools. Moreover, healthcare providers and insurers can benefit from partnering with or utilizing the website to enhance their patient care and wellness programs, further solidifying the business viability of this venture.

**Target Specification and Characterization:**

The target audience for this diabetes prediction website includes individuals of varying age groups who are concerned about their health and want to take pre-emptive actions to mitigate their risk of diabetes. The website caters to both tech-savvy users who are comfortable with online platforms and those who may be less familiar with digital tools but are motivated to explore health-related solutions. The platform's user-friendly design ensures that both segments of the audience can effortlessly input their health information and receive meaningful predictions. Through the online web portal, we can take user input and predict diabetes of the user with the help of ML algorithm. This web app can also recommend hospitals and doctors of high grade where patients can get quality treatment as well. Medicines can be brought to the home of the patients through online order in our website. So our web-app is the all-in-one solution to this emerging healthcare problem.

**Benchmarking Alternate Products:**

One of the primary alternate products in the field of diabetes prediction is traditional medical diagnosis conducted by healthcare professionals. This approach involves clinical assessments, physical examinations, and laboratory tests, which provide accurate results but often require significant time and resources. In contrast, the diabetes prediction website aims to offer a convenient and rapid method for individuals to assess their risk of diabetes. However, accuracy remains a key concern when benchmarking against traditional medical diagnosis. A thorough evaluation should address the website's predictive accuracy, sensitivity, specificity, and the ability to detect different types of diabetes.

Machine learning algorithms and predictive models for diabetes risk assessment also constitute alternate products. These models leverage extensive datasets and advanced analytics techniques to make predictions based on an individual's demographic, lifestyle, and health-related information. Benchmarking the website against these models involves comparing its predictive performance, ease of use, and accessibility. It's essential to assess whether the website's predictions align with established machine learning models in terms of accuracy and reliability. Furthermore, considerations should be given to the interpretability of results, as complex models might lack the transparency that a user-friendly website can provide.

Other alternate products include mobile applications focused on diabetes management and risk assessment. These apps often offer features like tracking blood glucose levels, monitoring dietary habits, and providing personalized recommendations. When benchmarking against these apps, the diabetes prediction website's usability, user interface, and integration of relevant data sources should be evaluated. The website's advantage lies in its singular focus on risk prediction, potentially offering a more streamlined and straightforward experience compared to multipurpose apps.

**Applicable Patents:**

Designing a Model to Detect Diabetes using Machine Learning – Inventor ( [Shashi Bhushan](https://patents.google.com/patent/AU2021103883A4/en), [Sunil Gupta](https://patents.google.com/patent/AU2021103883A4/en), [Afshan Hassan](https://patents.google.com/patent/AU2021103883A4/en), [Pariza Kamboj](https://patents.google.com/?inventor=Pariza+Kamboj), [Harish Kundra](https://patents.google.com/?inventor=Harish+Kundra), [Birajashis Pattnaik](https://patents.google.com/patent/AU2021103883A4/en), [Devendra Prasad](https://patents.google.com/patent/AU2021103883A4/en), [Parveen Sharma](https://patents.google.com/?inventor=Parveen+Sharma), [Basant Verma](https://patents.google.com/?inventor=Basant+Verma) ) was publishes in 2021.

**Applicable Regulations:**

Data Protection and Privacy: It is important to inform users about the data being collected, obtain their consent, and implement appropriate security measures to protect user information.

Medical Regulations: Medical Council of India (MCI) provides guidelines for telemedicine practices, which may be relevant to online healthcare services. It is important to ensure that the product comply with relevant medical regulation.

Data Localization: India has proposed regulations related to data localization, which may require certain data to be stored and processed within the country.

Intellectual Property Rights: It is important to protect our intellectual property rights through patents, copyrights, or other appropriate measures.

**Applicable Constraints:**

Privacy and Data Protection: It must be ensured that the data collected from users is handled securely and in compliance with applicable privacy laws.

Ethical Considerations: Measures should be taken to mitigate potential biases and it should be ensured that the model is fair and reliable across different demographic groups.

Data Availability and Quality: The accuracy and reliability of a prediction model heavily depend on the quality and availability of the data used for training.

User Interface and User Experience: The prediction model should be integrated in a user-friendly manner, providing clear instructions and intuitive interactions to ensure a positive user experience.

Expertise and Resources: Building and maintaining a prediction model requires expertise in machine learning, data science, and web development. Consider the availability of skilled professionals and the necessary resources (e.g., time, budget, infrastructure) to develop, deploy, and maintain the model on your website.

**EXTERNAL SEARCH:**

Machine Learning for Diabetes: A Systematic Review.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6234853/>

Predictive yodelling in diabetes.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276449/>

Machine Learning Techniques for Diabetes Prediction: A Review.

<https://www.researchgate.net/publication/326475266_Machine_Learning_Techniques_for_Diabetes_Prediction_A_Review>

Diabetes Prediction Using Machine Learning and Deep Learning.

<https://towardsdatascience.com/diabetes-prediction-using-machine-learning-techniques-37c6ebe0c2c9>

A Comprehensive Guide to Understand and Implement Text Classification in Python.

<https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/>

An Overview of Machine Learning for Diabetes Diagnosis and Prediction.

<https://www.jmir.org/2020/1/e15449/>

How Machine Learning Can Improve Diabetes Prediction.

<https://healthitanalytics.com/news/how-machine-learning-can-improve-diabetes-prediction>

**CONCEPT GENERATION:**

The increasing prevalence of diabetes worldwide poses a significant challenge to public health. As healthcare systems look for innovative solutions to manage this epidemic, AI emerges as a promising avenue. It’s not just about disease prediction but also monitoring, personalized care, and proactive interventions. With the success stories of AI applications in other medical fields, the intuition is that similar advancements can be made in the realm of diabetes care.

In cities like Bangalore, known for both its IT prowess and rising lifestyle diseases, there are efforts to amalgamate AI with medical diagnosis and predictive healthcare. However, despite the advancements, many healthcare facilities and providers are yet to harness the full potential of AI. As diabetes is multifaceted and can result from various physiological and environmental factors, AI can indeed be instrumental in offering more precise and personalized care.

**CONCEPT DEVELOPMENT:**

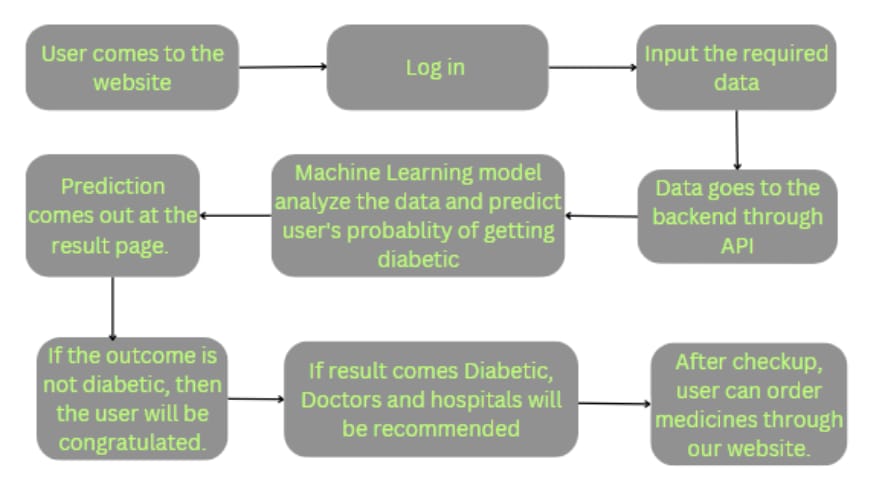
Before plunging into predictive analytics for diabetes, it’s imperative to comprehend the holistic environment of the patient – genetic factors, lifestyle choices, dietary habits, and even geographical implications.

With this contextual understanding, data collection becomes the next pivotal step. This should encompass not only medical histories but also daily activities, dietary patterns, and perhaps even real-time glucose monitoring for richer insights. Following data accumulation, Exploratory Data Analysis (EDA) is vital to discern patterns, trends, and anomalies within the dataset. Visualization tools can greatly aid in this process, enabling clearer interpretations of patient clusters, risk factors, and more.

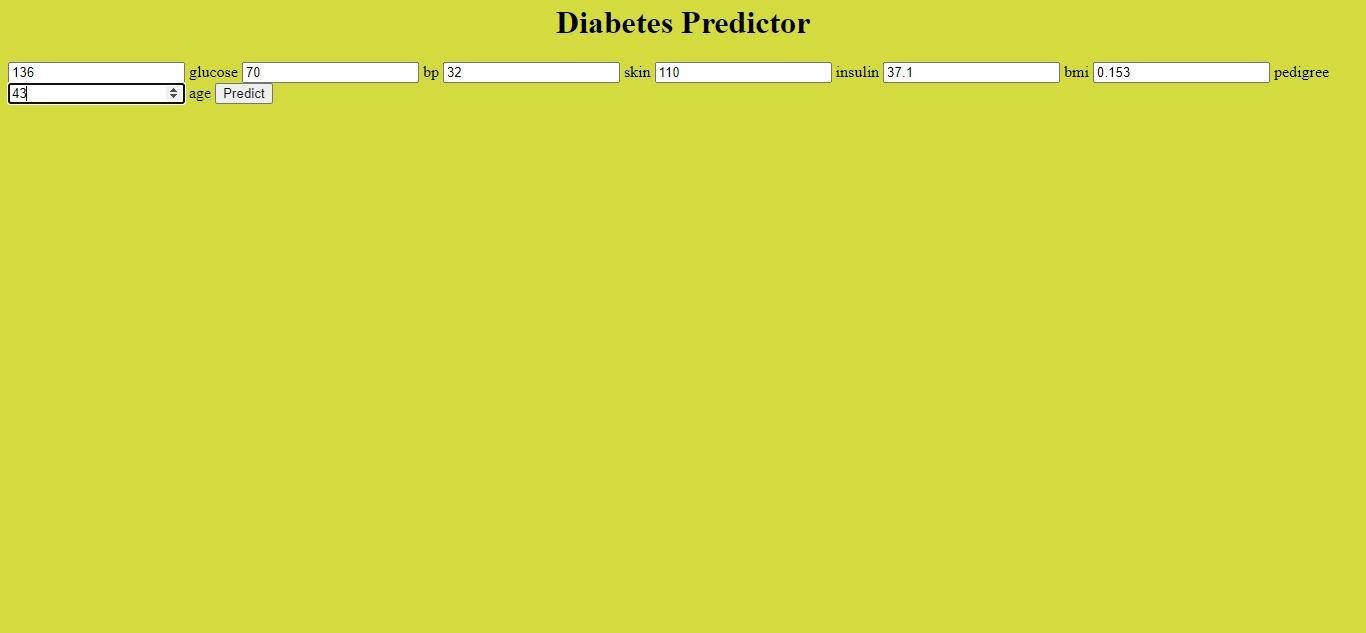
Building on this foundation, the yodelling phase can commence. Preliminary models, perhaps based on logistic regression given the binary nature of diabetes prediction (yes/no), can be developed. Evaluation metrics such as the area under the ROC curve, precision, recall, and F1 score can provide insights into model performance. Based on the performance, model refinement can be done, potentially incorporating more complex algorithms or ensemble methods to improve prediction accuracy. With the continuous influx of data, the models can be periodically retrained, ensuring they adapt and remain relevant as medical knowledge and patient data evolve.

**FINAL PRODUCT PROTOTYPE:**

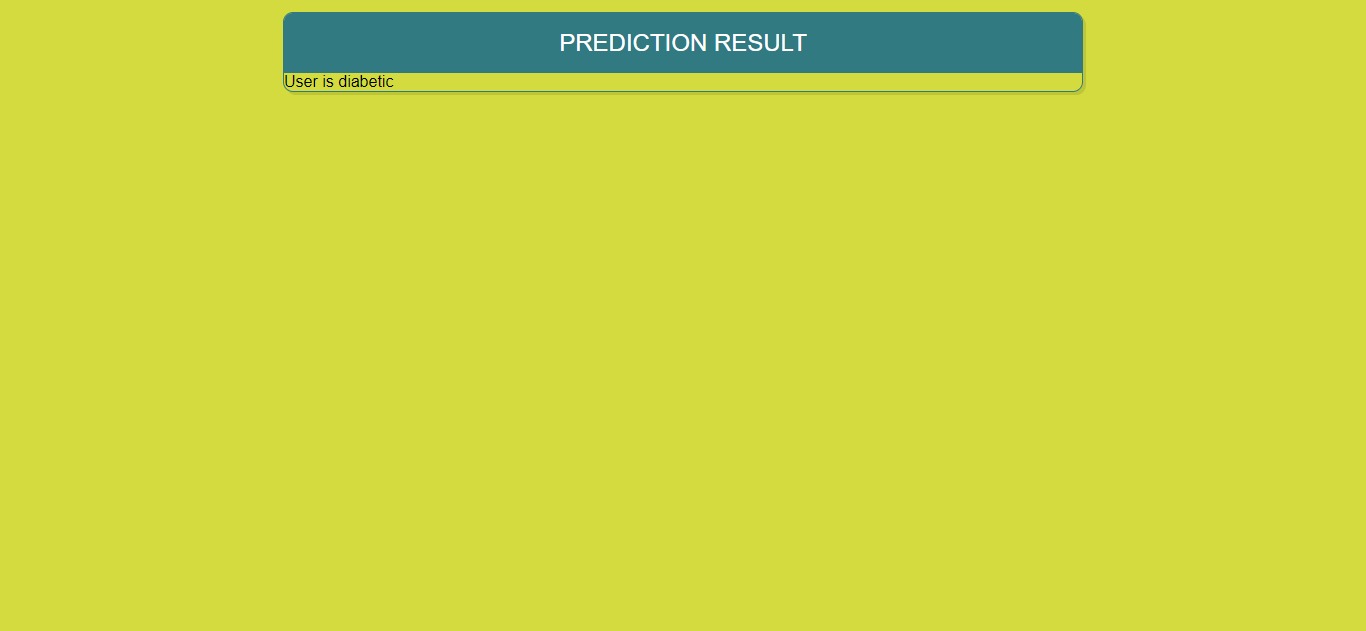
Flow chart:



Home page of the website will have a look like this:



Result page of the website will have a look like this:



We developed a basic website using Flask and deployed an ML model.

Regular fitting of the final model to the incoming data is necessary, and we should constantly be on the lookout for new features that might be affecting predictions.

**Back-end**:

🡪 This calls for the gathering of data, pre-processing, and integration of the model with the web application.🡪 The customer's consent should also be obtained before collecting and storing any of their entered data.

**Front-end**:

🡪The front end is important since it is the user interface that the customer will interact with.

🡪The web application could have mainly two pages. Input taking on the first page, then result comes on the second page.

🡪It needs to be very user-friendly; otherwise, users can enter incorrect information and the prediction will be completely wrong.

**Product Details:**

The product works through some steps, which are as follows:

1. User comes to the website and put the required input data after Log in.
2. Inputted data goes to the Machine Learning model through API.
3. Machine Learning model predicts diabetes of the user.
4. Predicted result comes out in the result web-page.
5. If the user gets predicted result as diabetic, an appointment with doctor will be booked and hospitals will be recommended.
6. Patients can order medicines for home-delivery after health checkup.

Required Technologies:

🡪Machine Learning classification algorithm

🡪 Python analytical libraries

🡪Flask

🡪HTML

🡪CSS

🡪Javascript

🡪AWS

Team required to develop the model:

🡪**Machine Learning Engineer:** Design, experiment and implement ML models.

🡪**Full-stack developer:** Design, test and implement applications.

🡪**Data Scientist:** Identify, undertake and analyze data.

🡪**Business Analyst:** Identify effective ways of boosting organizational efficiency.

🡪**DevOps Engineer:** Build, test and maintain infrastructure and tools, so that software can be developed and released.

**CODE IMPLEMENTATION:**

**GitHub links:**

1. <https://github.com/SpandanBandhu/Diabetes-Prediction-ML-Project>
2. <https://github.com/SpandanBandhu/Diabetes-Prediction-Website>

**Algorithm:** We are using Supervised Machine Learning model for this project implementation. A Classification algorithm is needed to classify our dataset into binary output (such as “yes” or “no”, “pass” or “fail” and so on). Decision Tree Algorithm has been implemented here and achieves 100% accuracy.

**Python Libraries used**:

🡪pandas: Pandas is defined as an open-source library that provides high-performance data manipulation in Python. Data analysis becomes easier with this python library.

🡪numpy: NumPy is used for mathematical function purpose. NumPy arrays are more memory efficient and compact than python lists.

🡪matplotlib: Matplotlib is a visualization library of python. We can graphically represent data and their relations through Matplotlib.

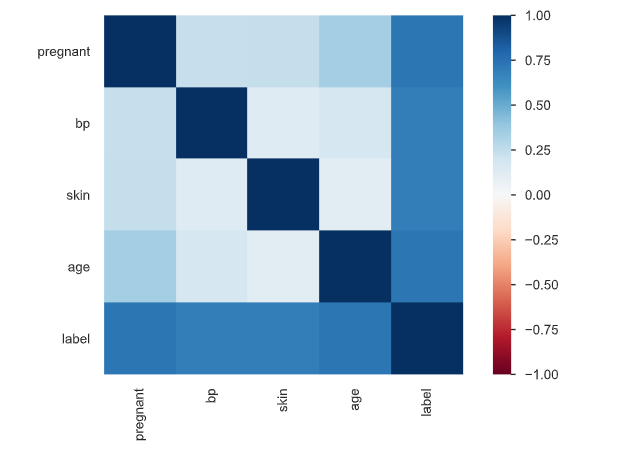
🡪sklearn: Sklearn is predictive analysis based python library. Data splitting, algorith implementation etc are done through Sklearn.

🡪flask: Python is used to create web applications with Flask, which is developed on Werkzeug and Jinja2. There are benefits to adopting the Flask framework, including a built-in development server and a quick debugger.

**EDA implementation:**

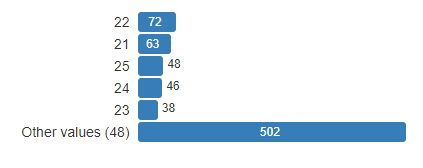






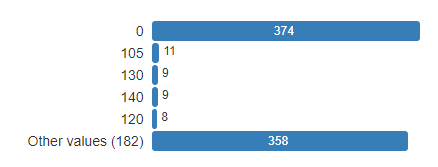
The feature ‘Pregnant’ is not affecting the output. So, we can remove the feature from our dataset.

**Age:**



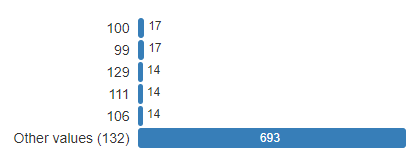
Here we can see young people of 21 – 23 ages are falling to the Diabetes disease.

**Insulin:**



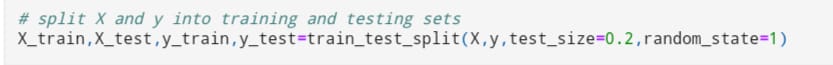
Insulin takes an important part in people’s chance of getting diabetes.

**Glucose:**

****

A significant factor in determining a person's risk of developing diabetes is glucose.

**Splitting Training data and Testing data:**



**Object creation of the model and checking accuracy:**



**Flask implementation at Backend:**



After getting input from the users, Machine Learning models comes into the action. Flask works as an API and fetches the inputs from home page to ML model. Model predicts output and the result is displayed at final page.

**Business Model:**

In this part of the report, we will look at the business model suggested for the idea presented earlier. There are many business models available but we have chosen the best suitable model for our dataset.

**Step1:**

We meet the user through online interface. As our model will be embedded in the website, so its benefit will gain maximum reach. The hassle of physical presence for diabetics check will be nullified.

**Step2:**

To give the personalized experience to all the user we have arranged the login system. This will also help the users to keep their trail of their diabetic checks.

**Step3:**

For diabetics check we need certain required details from the user like glucose level, bp, insulin, pregnancy check etc. We will take all these details using our interface.

**Step4:**

From the user interface these attributes will be passed to the backend through API.

And in the backend, we have kept our trained model which will show predictions based on those collected information from the respective user.

**Step 5:**

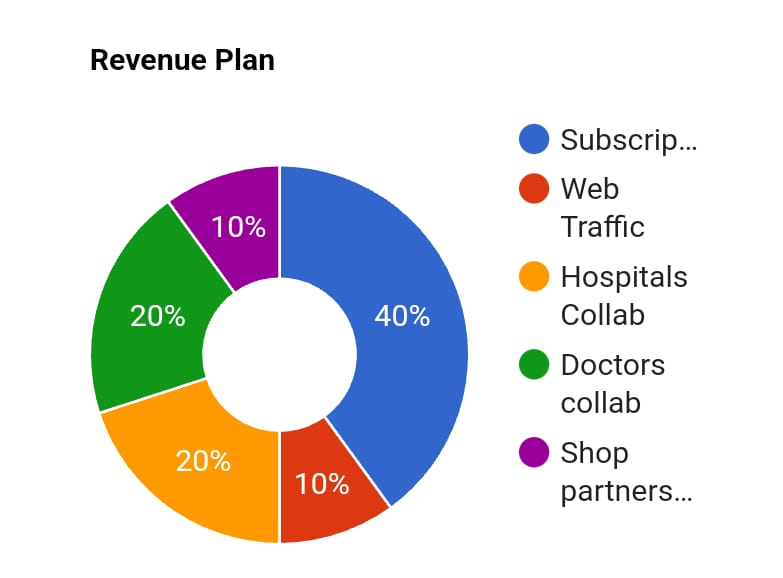
The prediction will be reflected in the webpage. If the prediction says that the user in not diabetic then he will be congratulated.

**Step 6:**

If prediction says that he/she is diabetic the proper recommendation of various hospitals and doctors will be shown, so that user can take proper steps to improve his/her health.

User will also be guided for their medication which are highly recommended by the prestigious doctors.

Now, For the financial part. We have 4 ways through which we generate our revenue and here they are:



**(i) Subscription:** People can use our product after subscribing to our website. Through website subscription they can get access to facilities like diabetes test, appointment of doctors, recommendation of hospitals, home delivery of medicines. It will make most of our revenue.

**(ii) Web Traffic:** We can advertise other companies through our website. Website traffic will bring bucks in this case.

**(iii) Collaboration with Hospitals**: Collaboration with hospitals can make us a good amount of profit. We will refer those hospitals in our website for treatment purpose in exchange of funds.

**(iv) Association with Doctors:** Doctors are one of the main pillars of our business model. We can charge an amount from the doctors whom we will be assigning for the patients.

**(v) Partnership with medical shops:** While undergoing treatment, patients can order medicines for home-delivery through our websites. Medicines will be provided with the help of local medical shops and a percentage of profit will be taken away from those shops.

**🡪Market Analysis:** The medical industry is in urgent need of such kind of automation. As we all know for any kind of surgery or serious operations the level of sugar should be at desired level. It becomes very difficult for any patient to keep a regular check of their diabetic level. As level of diabetics is unpredictable so it become crucial to keep it in regular check. But for regular check the physical presence at various clinics and leads to lots of chaos. To make this hassle free we came up this idea of developing diabetes check. This will reduce the hassle of regular check-up. Model will be train over huge number of records, so the accuracy of the model will be high and precise.

Since this is an exquisite combination of ML with medical problem so this implantation is new and attractive. Therefore, there is a huge opportunity for those who enter the market first.

**🡪Operating Plan:** The important part of our operation is to have ML/DS engineers with a good amount of knowledge about the industry. The product developing team’s size should be 3 to 4 where one of the members must be a full stack web developer and the remaining members must be ML engineers. It would be beneficial if almost all the ML engineers had knowledge about the industry.

The time for developing the product must be decided after a meeting with the client and the team developing the product. Having a clear idea about the deadline is a must and based on that the team can accelerate certain parts of the developing process.

When we start providing this service, we must set a low price because initially we need the maximum reach. Once we have successfully implemented our model for the first client and depending on our model’s performance, we can take stock about our position and decide on the pricing. Pricing should also be based on the type of hotel we offer our service to.

**🡪Marketing Plan:** For marketing our product, first we should collect the details of various hospitals and their responses over several diabetic’s scenarios, how many doctors are specialist and how much reputed the hostel is? These questions will decide the order of recommendation list which will be shown to the user. Since we will also recommend the medicine, a proper list of pharmaceutical shops should be arranged which supports home delivery, to give the user a highest possible comfort. Once the demand or reach of our website increases then we can increase our percentage profit from various hospitals and pharmaceutical shops.

**Financial Equation:**

As we all know the demand for medical services will always remain high and it will increase with time. Therefore, our product will have higher chance to boom. To decide the approx. salary of our team, let’s assume that to recommend the bigger hospitals we can charge around Rs. 15000 for three months and for pharmaceutical shops around Rs 10000 for 3 months. Once the customer base increases, we can either increase the price or reduce the duration for which our product will be available.

Let’s assume that the duration of developing the ML model takes about 1 to 3 weeks and the cost for producing the model is the salary of the members the team. Let there be two ML engineers and one full stack web developer. Let the salary of the ML engineers be ‘ml’

18

and the full stack web developer be ‘fs’. So, the total cost c = 2\*ml + fs. So, the profit or financial equation will look like this

**y = 25000\*x(t) – (2\*ml + fs)**

Here x(t) is a function that represents the growth of the customer base and y is the profit.

**Conclusion:**

Diabetes is not merely a disease but a burgeoning global health crisis. With millions affected and many more at risk, early detection becomes a matter of paramount importance. In many parts of the world, including several regions in India, there remains a discernible gap between medical need and available healthcare services. The disparity in healthcare quality, coupled with challenges in timely diagnosis, can lead to complications, reduced patient trust, and, at times, fatal outcomes.

Integrating machine learning into the diabetes prediction landscape promises a paradigm shift. By facilitating early detection, providing personalized risk assessments, and continuously evolving through data, AI-driven tools can be game changers. The solutions we are developing not only aim to bolster accuracy but also aspire to make the diagnostic journey more patient-centric, transparent, and efficient

As we stand on the precipice of a new era in healthcare, it’s conceivable that machine learning will soon become an indispensable ally in our fight against diabetes, transforming prognosis, management, and overall patient care.