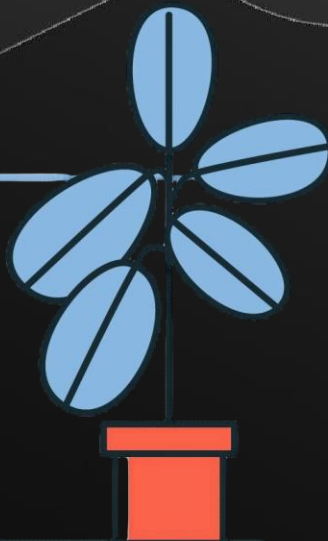




STARK HEALTH CLINIC – DIABETES PREDICTION PROJECT

Leveraging machine learning to predict diabetes onset and empower proactive healthcare interventions.

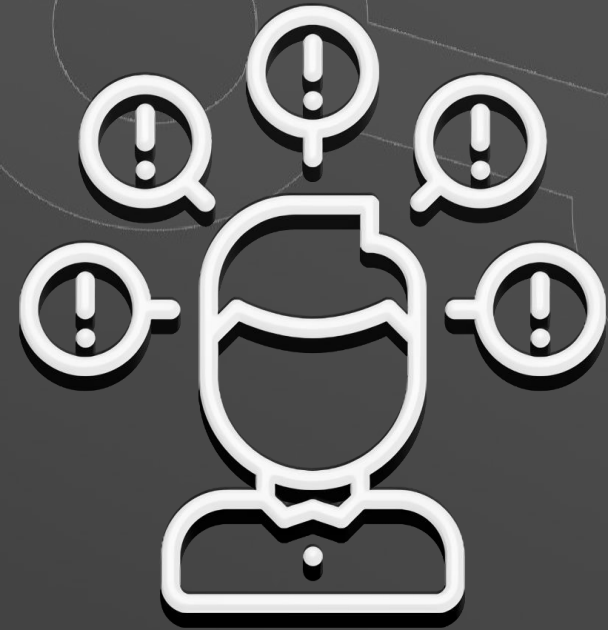


BUSINESS INTRODUCTION

Stark Health Clinic is a leading healthcare provider that leverages technology and predictive modeling to enhance its operations. By integrating machine learning into its systems, the clinic identifies diseases early, improving patient outcomes and resource allocation.

PROBLEM STATEMENT

Diabetes poses significant health risks to Stark Health's patients, and also financial challenges. Current methods for early detection at Stark Health Clinic lack precision, leading to missed opportunities for timely interventions.





RATIONALE OF PROJECT

By accurately predicting diabetes risk through advanced machine learning, Stark Health Clinic can improve patient care, reduce long-term costs, and take a proactive role in combating diabetes.

OBJECTIVE OF THE PROJECT

Stark Health Clinic aims to develop a robust diabetes prediction model to accurately identify individuals at risk of developing diabetes. You have been approached as a Data Scientist to lead this project, utilizing advanced machine learning techniques on patient data.

The goal is to predict the likelihood of diabetes onset, allowing for timely and targeted preventive measures. This initiative will empower Stark Health to enhance patient outcomes, reduce the burden on healthcare resources, and play a proactive role in combating diabetes.

TECH STACK

- Python
- NumPy
- Pandas
- Matplotlib/Seaborn
- SciKit Learn
- Anaconda & Jupyter Notebook



ANACONDA®



PROJECT GUIDE

1. PROBLEM DEFINITION: clearly articulate the problem that is to be solved with your data mining. How will the company benefit from your solution?

2. PERFORM EXPLORATORY DATA ANALYSIS IN PYTHON:

- a) Visualize relationships between the target and some key features
- b) Explore correlations
- c) Conduct univariate, bivariate, and multivariate analysis as much as is feasible

3. PERFORM FEATURE ENGINEERING:

- a) Encoding categorical variables
- b) Create new features from existing features where necessary, depending on insights from your EDA

4. MODEL SELECTION, TRAINING, AND VALIDATION:

- a) Train and test at least 3 supervised learning models

5. MODEL EVALUATION:

- a) Analyze the results of your trained models
- b) What metrics are most important for the problem? Should the business be more concerned with better results on false negatives or true positives?

PROJECT COMPLETION

1. SUBMISSION:

- a) Publish your Jupyter Notebook to your GitHub profile and also submit it on the Google Classroom.
- b) In the readme file on GitHub, include a description of the project, summarize the steps you took and document the challenges you faced.
- c) share the link with your instructor.

2. PRESENTATION:

- a) While you work on your project, prepare a PowerPoint presentation of your project
- b) Apply the S.T.A.R Approach in preparing your slides and documenting the outcome of your analysis
- c) You will have 5 minutes to present your work to the stakeholders using your slides, and will be receiving feedback on your project and presentation

NB: You are expected to submit both your Jupyter Notebook, and your presentation slides on the google classroom