

Problem Statement

Knowledge Representation and Insights Generation from Structured Datasets

Develop an AI-based solution to process and analyze structured datasets, effectively representing knowledge and generating meaningful insights to aid decision-making. The solution must include data pre-processing, knowledge representation, pattern identification, insight generation, scalability, and a user-friendly interface for easy interaction and understanding of insights.

Team : Fab Four

Vallu Sai Gopal (RA2211026010506) SRM IST

Sanjanya (RA2211026010479) SRM IST

Harash Bayana(RA2211026010532) SRM IST

Hashil Jain (RA2211026010513) SRM IST

Unique Idea Brief (Solution)

We developed an AI-based solution using a sample dataset named "10-Year Heart Attack Risk Prediction." Our solution pre-processes the dataset and generates insightful analysis, presented through a user-friendly web interface. Users can easily explore the dataset's analysis and input their custom data for personalized insight generation, ensuring a straightforward and engaging experience.

Features Offered

Data Pre-processing and Cleaning: We have performed extensive data cleaning and pre-processing on a dataset called "10-Year Heart Attack Risk Prediction." This includes handling missing values, normalizing data, and ensuring the dataset is ready for analysis.

Insight Generation: We have generated several insights from the dataset, identifying key patterns and trends that can help in understanding the data better. These insights are displayed clearly on the website.

ML Model Comparison: We tested various machine learning models on the dataset to determine which model performs best for heart attack risk prediction. The results of these comparisons are available for users to review.

Downloadable Analysis: The detailed analysis and pre-processing steps have been compiled into a Colab notebook, which users can download directly from the website for further exploration.

Custom Input:

Users can enter their own values for specific variables.

Based on these inputs, the platform leverages the best-performing machine learning model to generate predictions.

Custom Dataset Upload:

Users can upload their own dataset containing variables similar to those supported by the platform.

The uploaded dataset undergoes the same pre-processing, analysis, and insights generation process as the platform's internal data.

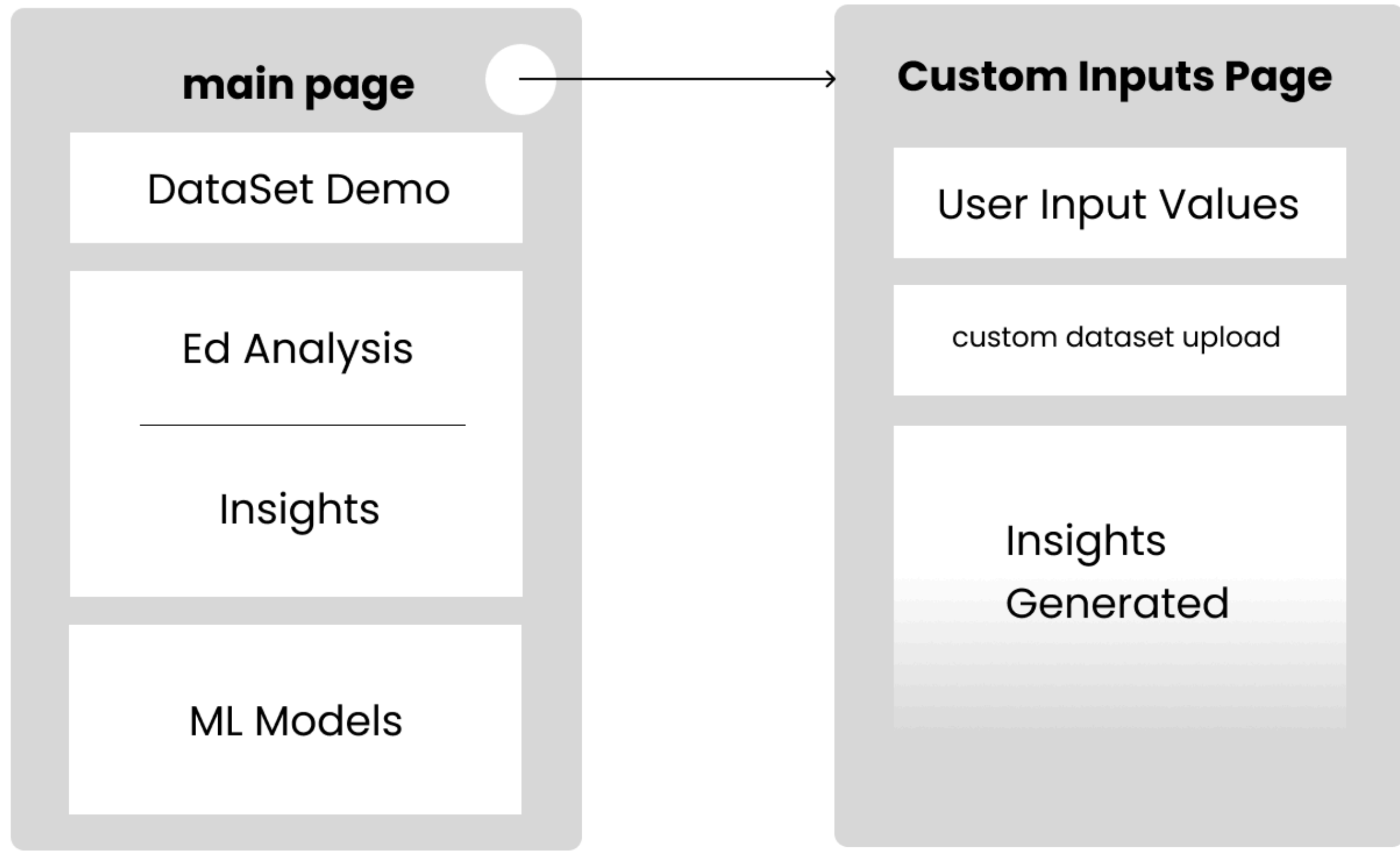
This allows users to get customized insights from their own data.

Process Flow

- **Homepage:** Introduce the project focused on predicting 10-year heart attack risk using machine learning. Highlight the importance of the dataset and the goal of providing personalized risk assessments.
- **EDA Page:** Provide an in-depth analysis of the dataset, including cleaning and preprocessing steps such as handling missing data and encoding variables. Showcase visualizations and key insights discovered about factors influencing heart attack risk.
- **Model Comparison:** Compare various machine learning models (e.g., logistic regression, random forest) based on performance metrics like accuracy and interpretability. Select the most effective model for predicting heart attack risk.

- **Prediction Page:** Allow users to input their own data (e.g., age, cholesterol levels, BMI) or upload a dataset. Utilize a Flask-based logistic regression model to process inputs and generate personalized risk predictions.
- Display predictions for user inputs or uploaded datasets. Provide insights into risk factors identified through the analysis and suggest potential actions based on the predicted risk levels.

Architecture Diagram



Technologies used

Machine Learning (ML) Part:

- **Python:** Utilized for data analysis, model training, and prediction. Libraries like Pandas for data manipulation, **Scikit-learn for machine learning models (e.g., logistic regression)**, and **Matplotlib/Seaborn** for data visualization.
- **Flask:** Developed a **RESTful API** to serve machine learning models. Used to integrate the logistic regression model for predicting heart attack risk based on user inputs.
- **Colab:** Used for **exploratory data analysis (EDA)** and initial model training. Colab notebooks were integrated into the website for transparency and reproducibility of analysis.

Website Building:

- **Svelte:** Frontend framework used for building the interactive user interface (UI) and managing user interactions.
- **Tailwind CSS:** Used for styling and layout of the website. Tailwind's utility-first approach facilitated quick prototyping and customization of UI components.
- **JavaScript :** Integrated to handle frontend logic and interactions, ensuring dynamic content updates and seamless user experience.

Team members and contribution:

Vallu Sai Gopal (Lead):

- Led the project with a focus on insight generation and building machine learning models.
- Conducted exploratory data analysis (EDA) using Colab notebooks.
- Contributed to the development of the Flask-based API for integrating machine learning models into the website.

Sanjanya:

- Specialized in frontend web development.
- Designed and implemented the user interface (UI) using Svelte and Tailwind CSS.
- Ensured the website's frontend was intuitive and responsive for user interactions.

Harshil Jain:

- Conducted detailed exploratory data analysis (EDA) to understand and preprocessing
- Contributed to frontend development, focusing on UI/UX improvements and data visualization integration.

Harsh Bayana:

- Played a key role in frontend development.
- Implemented frontend features, including user input forms and integration with backend services.
- Ensured smooth integration of the frontend with backend functionalities.

Conclusion

In conclusion, working on this project has been a valuable learning experience for our team. We gained insights into predictive modeling, data analysis, and web development. We are grateful to the Intel Unnati program for providing us with this opportunity to tackle such an important problem statement. This experience has strengthened our skills and passion for leveraging technology to address healthcare challenges.