COMMUNITY COMPANION CHALLENGE 2024

TEAM CODE:
DA24

COMMUNITY COMPANION CHALLENGE

2024

The Hidden Factors:
Exploring the Impact of Social Determinants on Health

Understanding The Problem Statement

- Lack of a Comprehensive Picture due to binary evaluation
- Surveys and Self Reporting and time consuming and are not reliable

HOW TO COMPENSATE

SDOH DATA

- Health Disparities
- Social Factor Affecting
- Healthcare Costs

GOAL

1. **RISK PROFILE**

   Generate risk profile based on patient data input

2. **ACTIONABLE INSIGHTS**

IMPORTANCE

FINE GRAINED ANALYSIS

We considered interdependencies between social determinant variables

QUICK ACTIONS

These actions are catered to every individual to provide effective support.

PREVENTIVE MEASURE

Solve problems before they become urgent and/or lead to other problems.

Dataset Insights

Problems

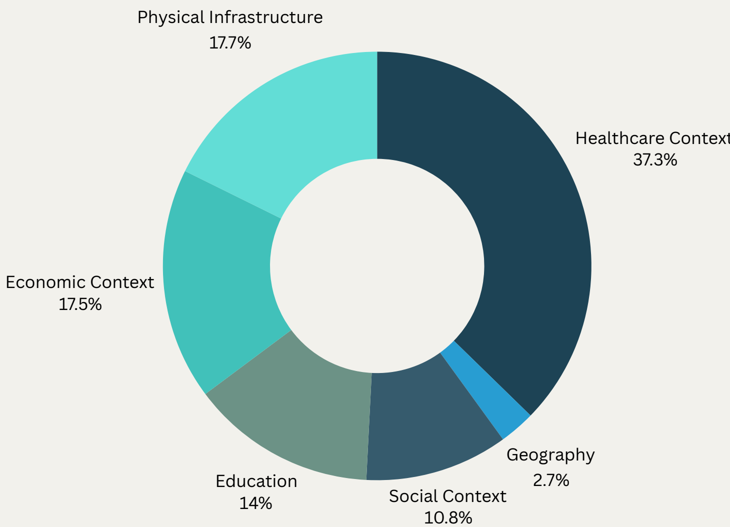
- Surplus of variables
- Numerous missing values
- Risk Identification and Assignment difficulty
- Large number of Zip codes
- Non uniform variables across counties, Zip codes and tracts

1405 Features

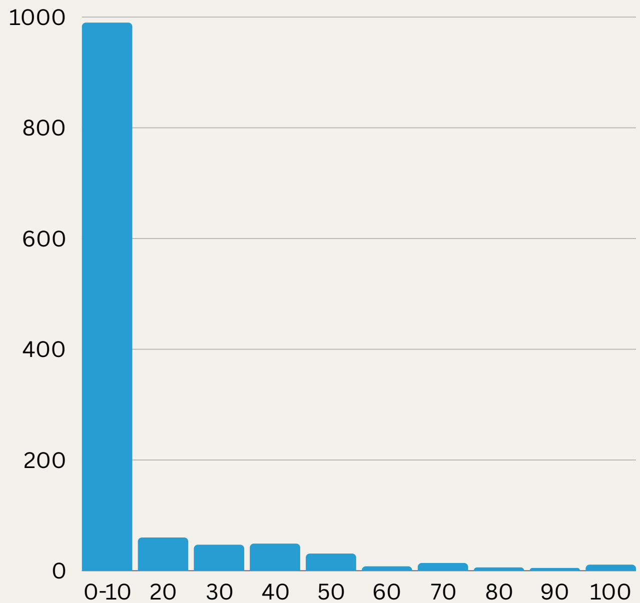
3K Counties

41K Zip Codes

85K Tracts



Distribution of available data for the different domains of SDOH.



Distribution of variables by their null value percentages.

COMMUNITY COMPANION CHALLENGE 2024

Dataset Preparation

1. Outliers Removal

Some SDOH variables exhibit outliers diverging from typical values

Outliers were detected and eliminated via boxplot analysis

2. Missing Values

Several variables exhibited a significant amount of missing data

Variables with missing data exceeding 50% were excluded

3. Unnormalised Values

SDOH variables displayed wide data range disparities.

Normalization was achieved through min-max scaling

4. Feature Name Discrepancy

Identical features across datasets had varying names

Features from diverse datasets were unified under consistent names

5. Absence of Primary Column

Datasets varied in primary columns, including tract, zip code, and county

Employed a zip code-census tract crosswalk dataset to link census tract and zip code

6. Missing values in 2020 data

SDOH variables exhibited numerous missing values for the 2020 dataset

Missing values were filled using the most recent available data from the 2018 or 2019 datasets

Initial Steps

Risk Identificaton

AHCM

Health Leads

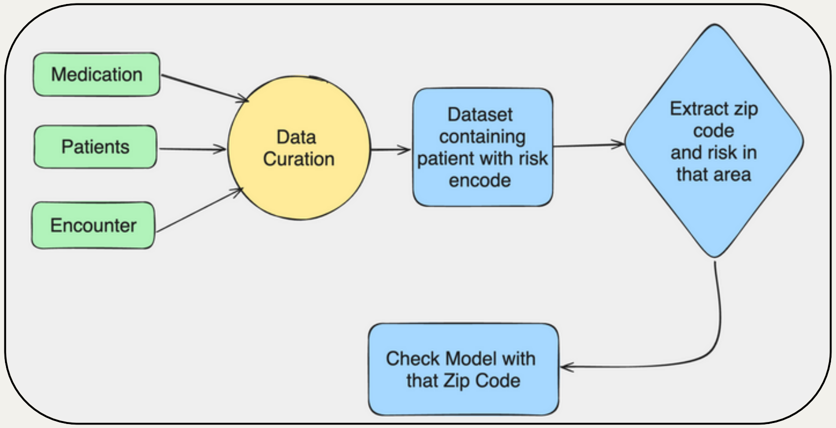
PRAPARE

Financial Risk

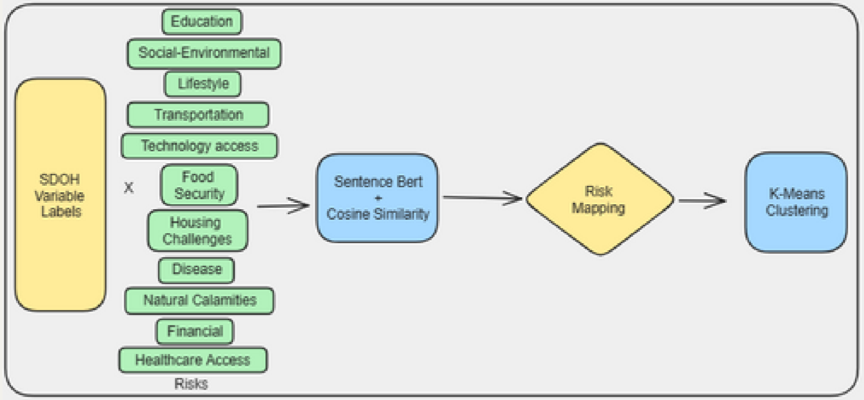
- In the last 12 months, have you needed to see a doctor, but could not because of cost?
- In the last 12 months, has the electric, gas, oil, or water company threatened to shut off your services in your home?
- How hard is it for you to pay for the very basics like food, housing, medical care, and heating?

11 RISKS IDENTIFIED

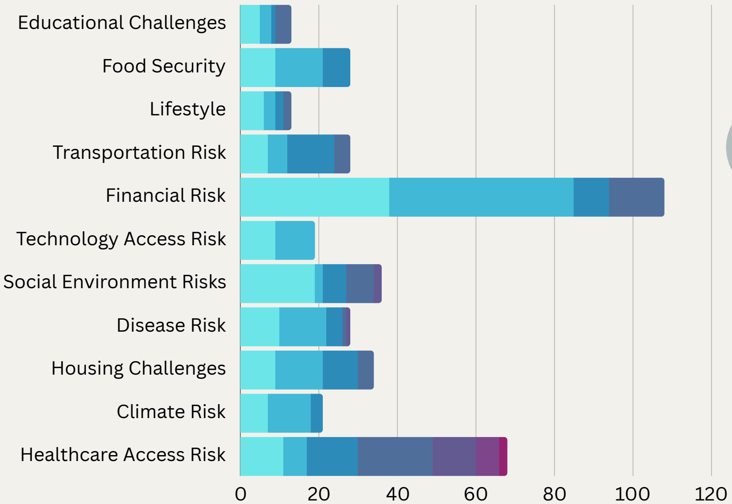
Feature to Risk Mapping



Input Parameter Mapping



Grouping Similar Clusters



Sentence BERT

+

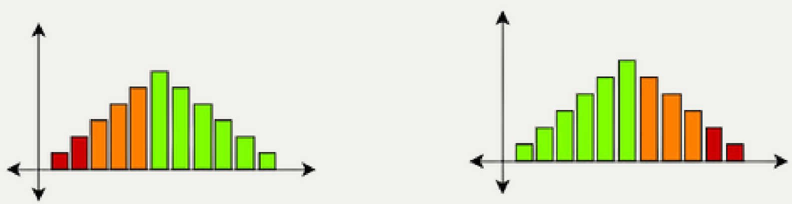
TF - IDF

COMMUNITY COMPANION CHALLENGE 2024

Towards the Scorecard

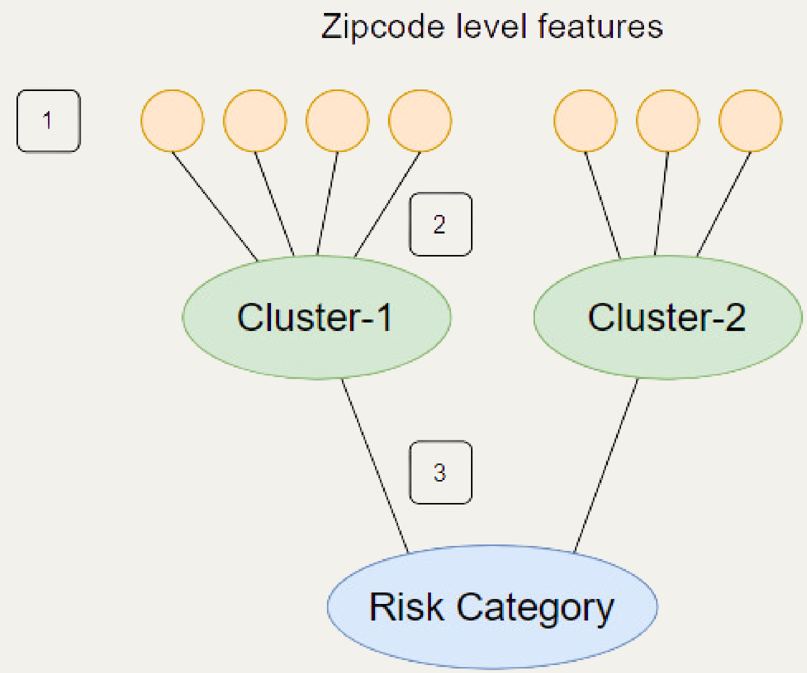
Distribution based thresholding for risk identification [1]

Depending on the direction of the influence of the feature on the corresponding risk identified, we set up percentile based threshold in the ordering high-mid-low or low-mid-high



Weighted Aggregation of Individual Risks on Cluster [2] and Risk Level [3]

The Individual Risks were added using a weighted combination based on whether the feature belonged to tract, zip or county level and then normalized. This was done at both cluster and risk level

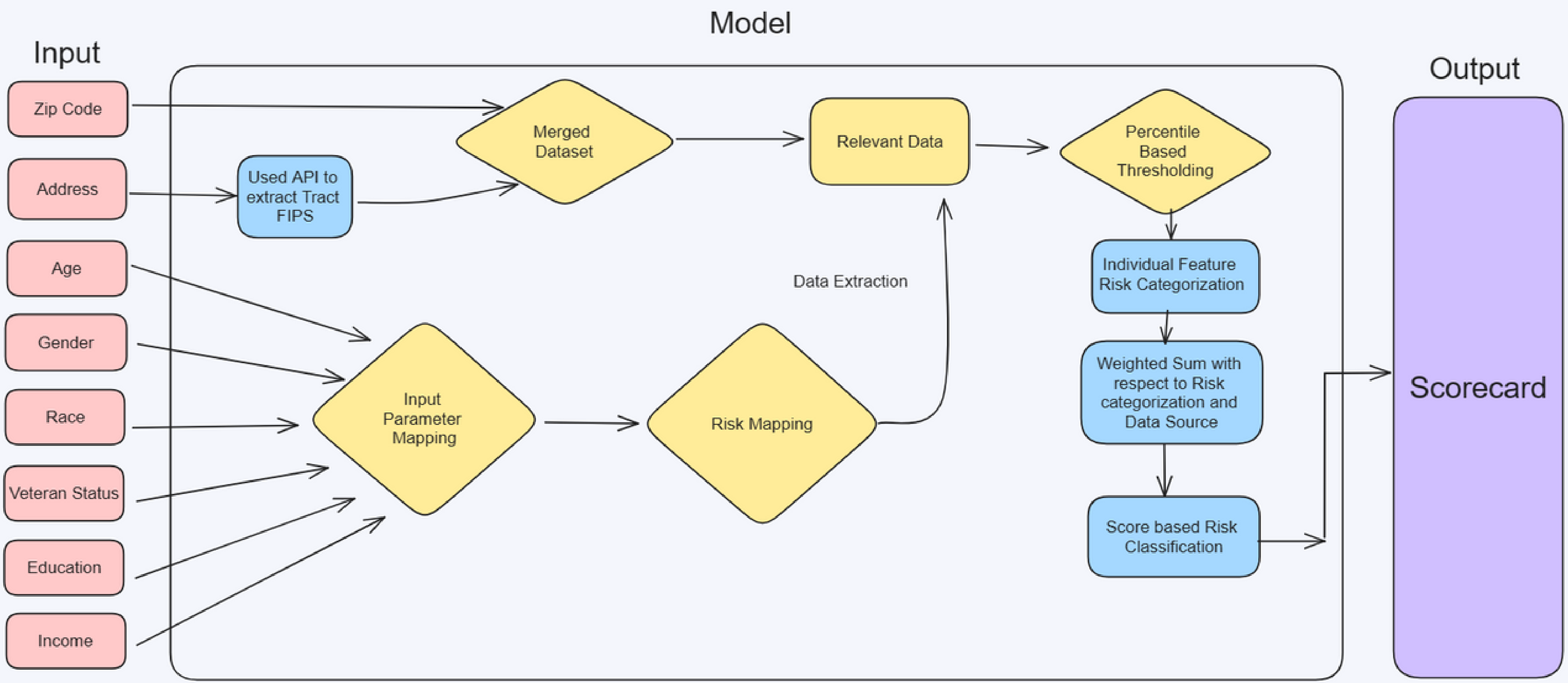


Additional Risk Categories

1. Disease Risk
2. Life Expectancy Risk

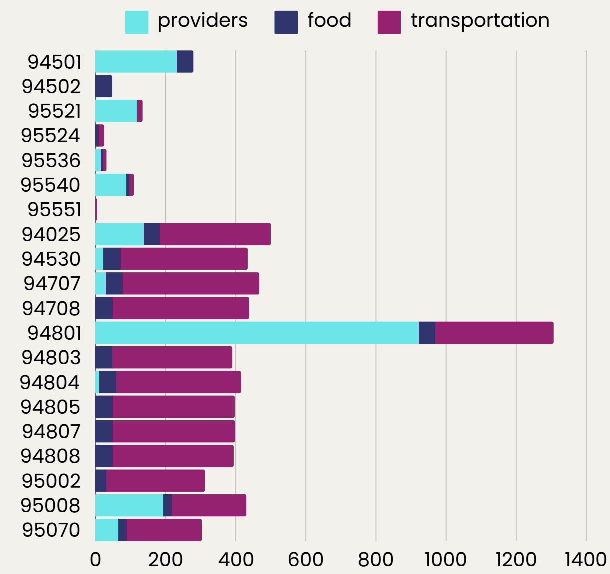
COMMUNITY COMPANION CHALLENGE 2024

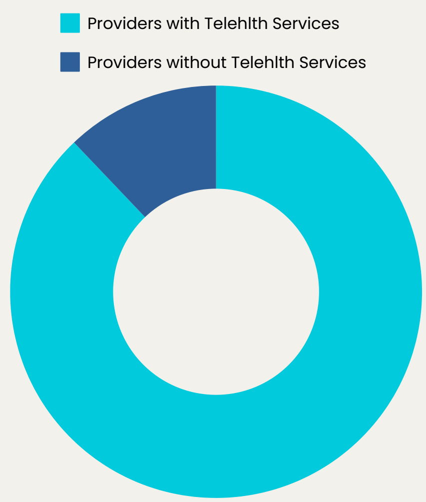
Model Pipeline

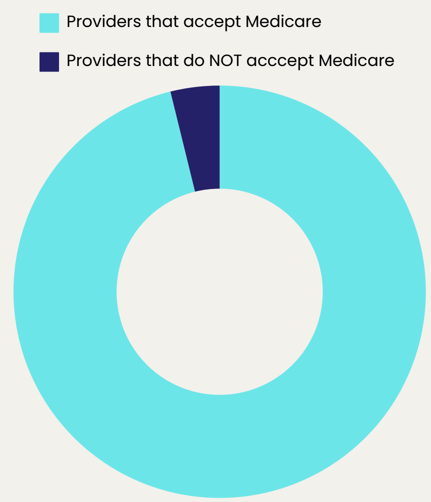


Actionable Insights

- **Fine grained insights**
  - We have suggested solutions based on specific sub-risks within each risk
- **Focusing on interactions**
  - Example: For Transportation risk, we have suggested tele-health services only if the technology access risk is low.
- **General as well as specific actions**
  - We have suggested nearest health service providers, hospitals, clinics based on needs for both patients and physicians







Curating a testing dataset - Synthea

Patient

- Contains patient's information such as age, race, gender, demographics, location, etc.
- Used to extract input parameters such as Zip code.

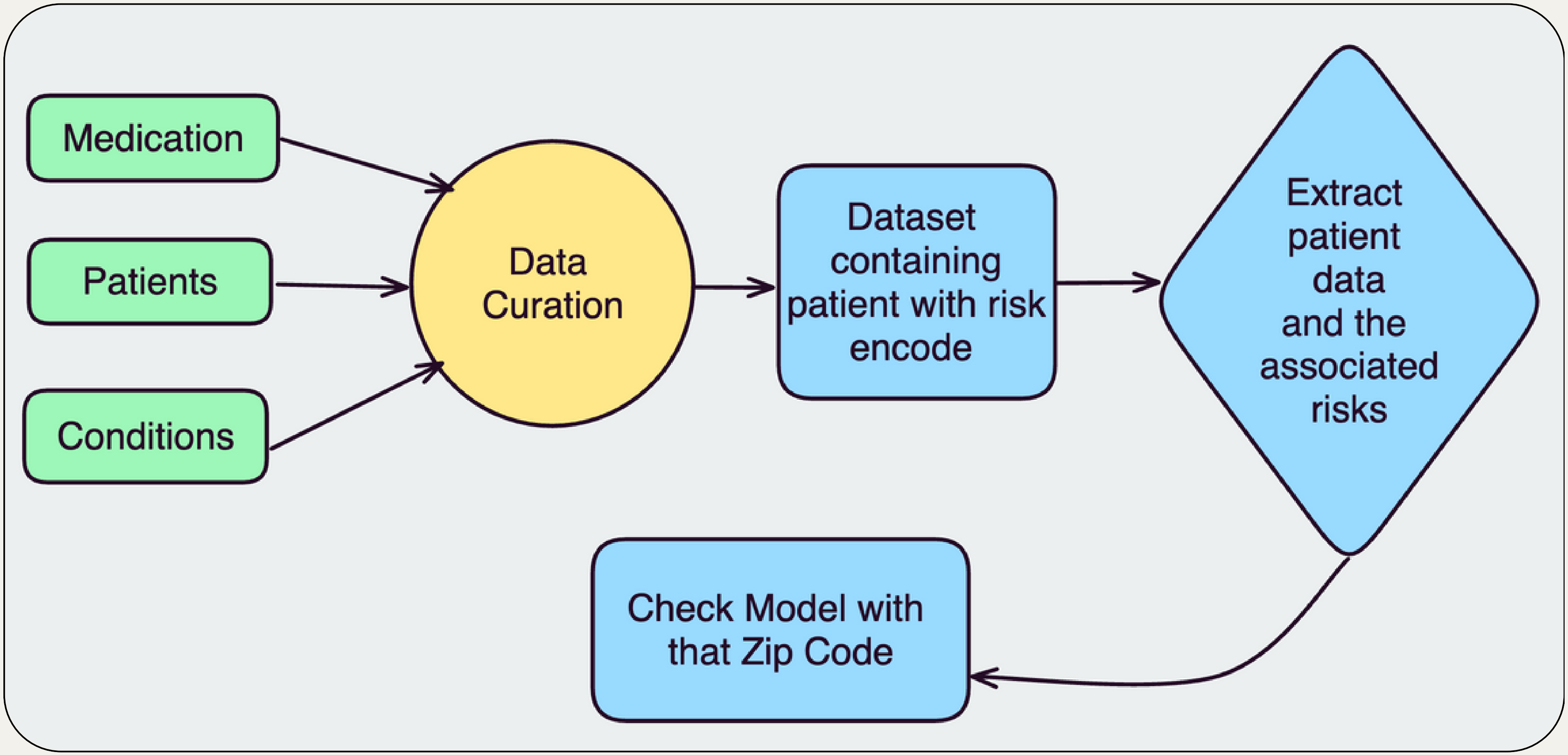
Condition

- Contains patient's health conditions as prescribed by the doctors.
- Used to associate various health conditions with risk identified by our model.

Medication

- Contains date and information of medication prescribed to the patient.
- Used to filter out relevant recent data (2018, 2019, 2020)

Curating a testing dataset - Synthea



Examples

Zip Code : 2151 | Age: 30 | M

Existing Risk : Lung Disease , Alcohol Usage

Risk Identified: Social Environmental Risk (Drug/Alcohol), Diseases RIsk (Lungs)

Actions : Provide Education and Awareness regarding Alcohol Abuse

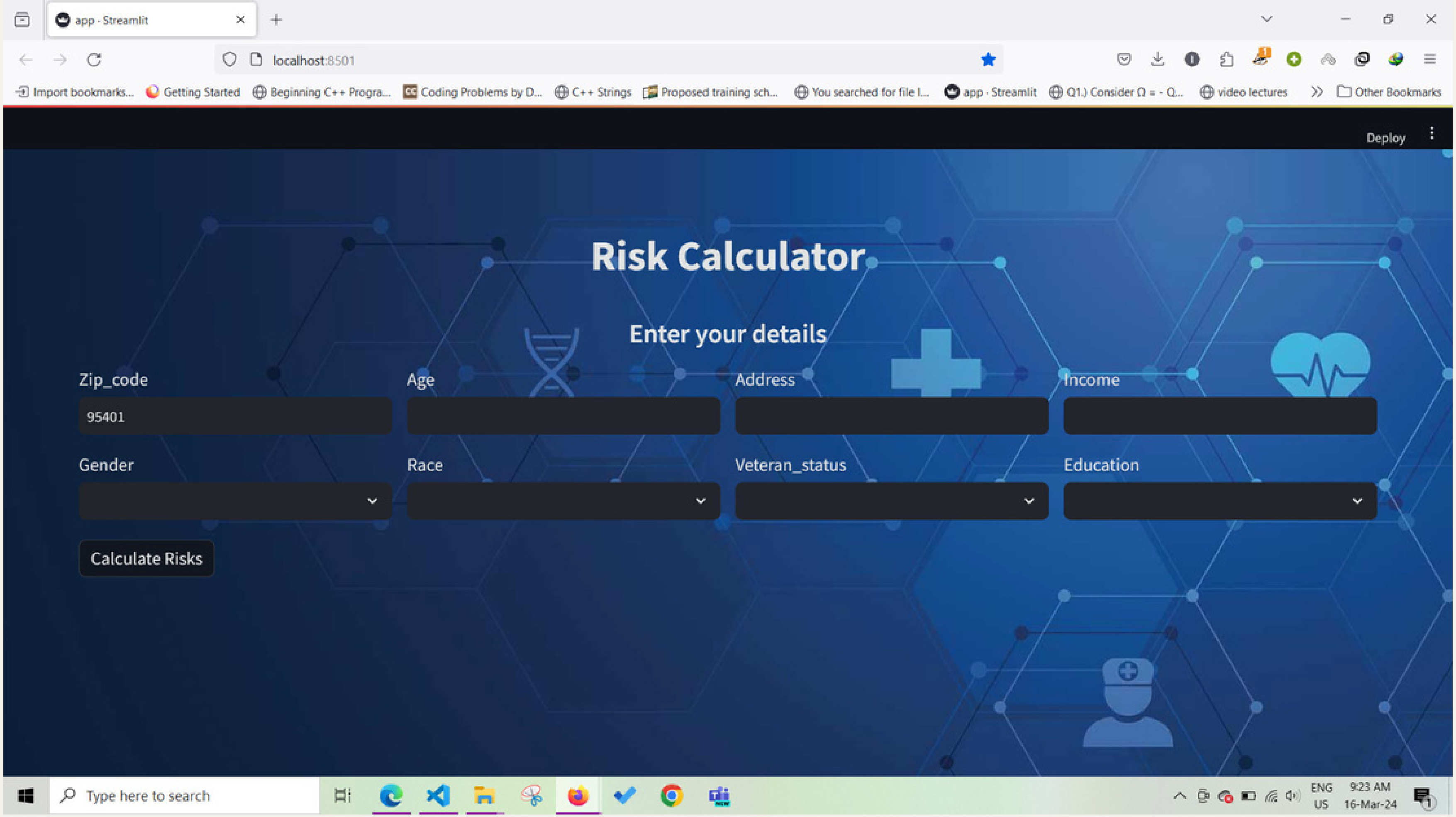
Zip Code : 2136 | Age: Unknown | F

Existing Risk : Transportation Risk, Mental Health

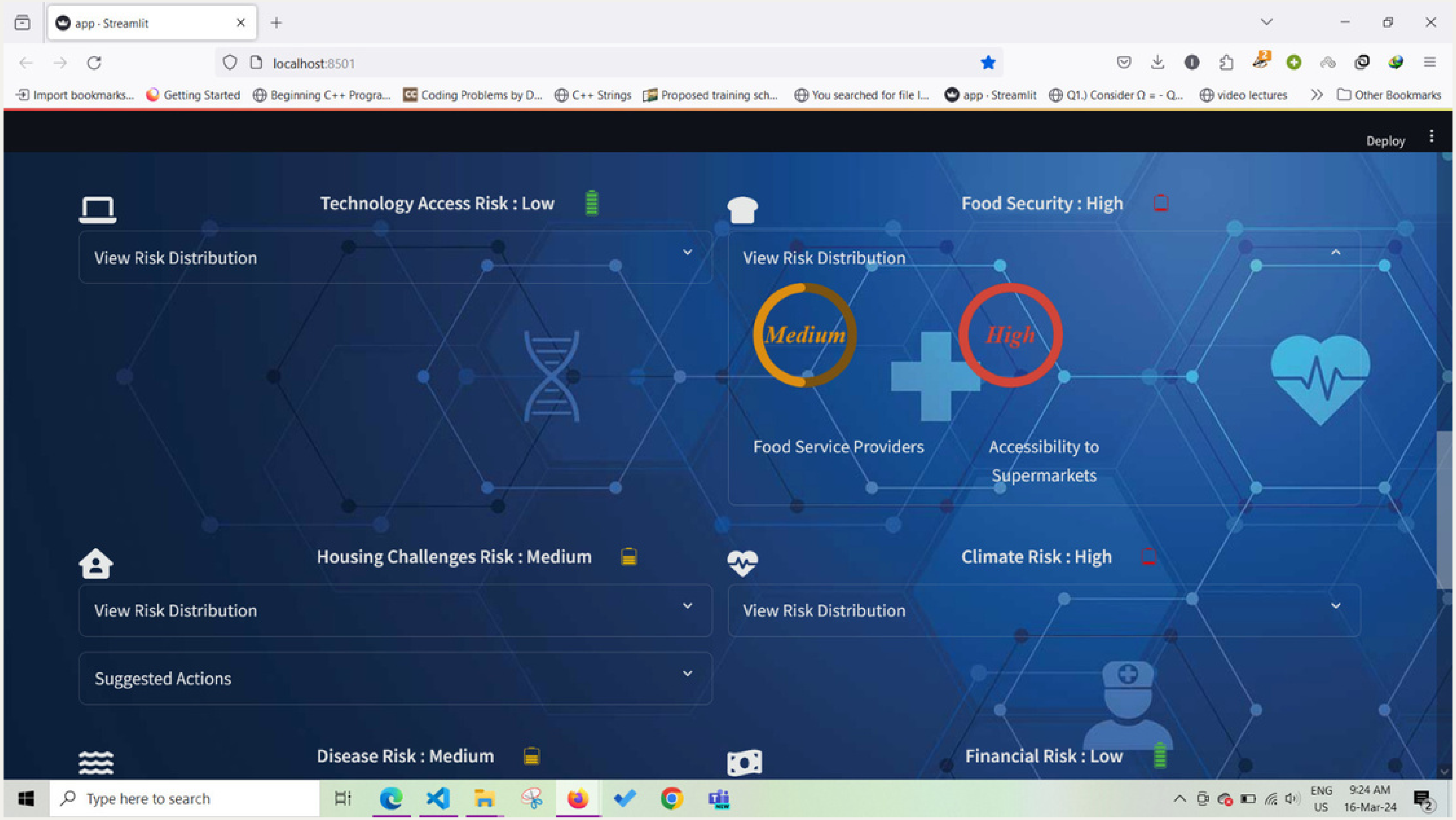
Risk Identified: Transportation Risk, Lifestyle Risk

Actions : Provide Transport Vouchers and Homecare provider

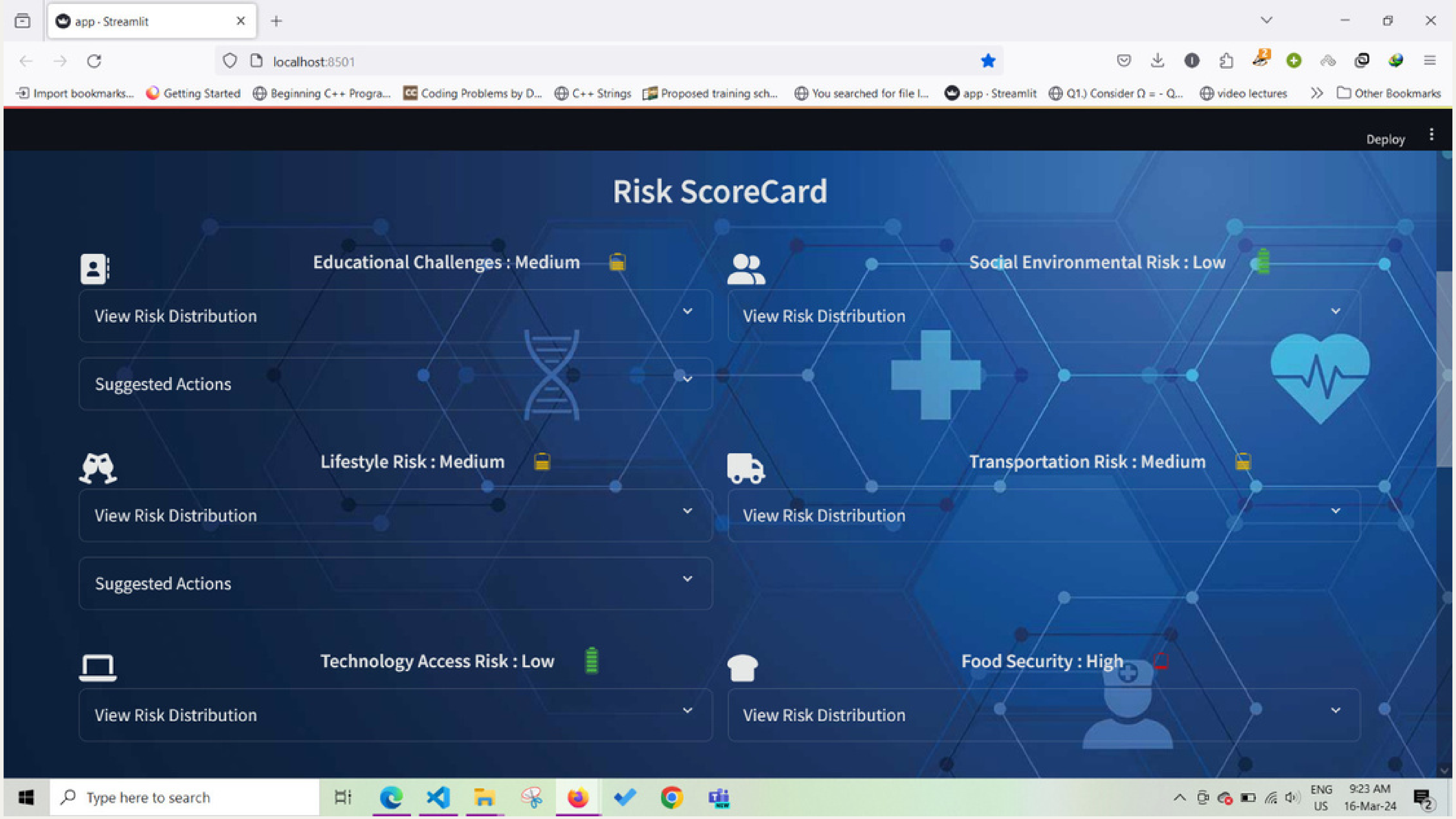
UI and UX


app - Streamlit
localhost:8501
Deploy
Risk Calculator
Enter your details
Zip_code
95401
Age
Address
Income
Gender
Race
Veteran_status
Education
Calculate Risks
Type here to search
9:23 AM
16-Mar-24

UI and UX


app · Streamlit
localhost:8501
Deploy
Technology Access Risk : Low
View Risk Distribution
Food Security : High
View Risk Distribution
Medium
High
Food Service Providers
Accessibility to Supermarkets
Housing Challenges Risk : Medium
View Risk Distribution
Suggested Actions
Climate Risk : High
View Risk Distribution
Disease Risk : Medium
Financial Risk : Low
Type here to search
9:24 AM
16-Mar-24

UI and UX


app · Streamlit
localhost:8501
Deploy
Risk ScoreCard
Educational Challenges : Medium
View Risk Distribution
Suggested Actions
Social Environmental Risk : Low
View Risk Distribution
Lifestyle Risk : Medium
View Risk Distribution
Suggested Actions
Transportation Risk : Medium
View Risk Distribution
Technology Access Risk : Low
View Risk Distribution
Food Security : High
View Risk Distribution
Type here to search
9:23 AM
16-Mar-24

COMMUNITY COMPANION CHALLENGE 2024

The India Challenge

1

What is different

- Disparities in access, the organization of healthcare services, and data quality
- Lesser technological advancements and infrasturcture developements
- Need for advance healthcare delivery in India

2

What do we deliver

- The predictive model employs automated clustering, risk detection, and thresholding processes
- Adaptable to diverse datasets, including the socioeconomic and healthcare variables critical for India
- Can address the unique challenges of the Indian healthcare

3

Future works

- Adjust the model's thresholds to align with India's specific healthcare needs
- Incorporate features into the model that reflect India's unique healthcare challenges

INDIA VS. US: HEALTHCARE INDUSTRY

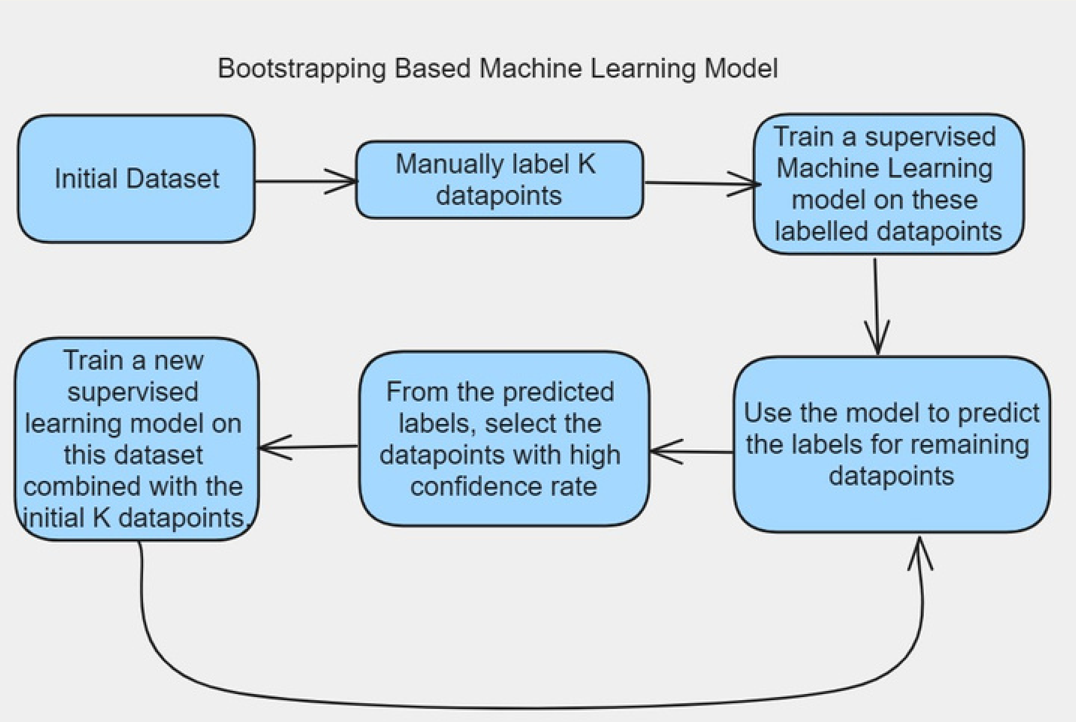
	INDIA	US
WHO HEALTHCARE RANKS	112	37
LIFE AT BIRTH EXPECTANCY	63 years for men and 66 years for women	76 years for men and 81 years for women.
PUBLIC HEALTH SCENARIO	spent about $40 per person annually	spent $8,500 per person annually
The entire GDP of India was $1.6 trillion then while the US health care spending alone was $2.6 trillion. In the US currently, per person healthcare expenditure is the highest in the world at an average of $10,345 per person.		
HEALTH SPENDS AS % OF GDP	The total expenditure on healthcare as percentage of GDP is just 4%,	It is 17%.
OUT OF THE POCKET EXPENDITURE	70% of the Indian population pays out of their own pocket for medical expenditures which is a staggering number compared to the US, the out of the pocket expenditure is much lower at 10-12%.	

Conclusion

- The model utilizes a comprehensive set of features to predict various risks, demonstrating its ability to handle complex datasets
- It efficiently identifies and categorizes each risk and associated sub-risks into high, medium, or low levels, offering targeted preventive actions to mitigate these risks
- By facilitating early intervention and promoting preventive care, the model empowers communities to proactively safeguard their health against various diseases, contributing not only to disease prevention but also encouraging healthier lifestyles overall

Future Work

- Explore model training enhancements through bootstrap methods, employing either machine learning or deep learning techniques for improved robustness and accuracy.



- Implement language model (LLM) fine-tuning and evaluate perplexity scores to assess output confidence and refine predictive performance.

THANK YOU

We invite any questions