Predictive Healthcare: Machine Learning for Optimized CEA Testing in Colorectal Cancer Patients

Team members:
 Ron Dickerson
 Mahima Joshi
 Janet Pan
Anuradha Ramachandran

Project Overview

- Background: Patients who have completed curative treatment for colorectal cancer (CRC) should be followed up for the development of recurrent disease and require surveillance for five years post-treatment

- Clinical guidelines: Suggest carcinoembryonic antigen (CEA) testing every three to six months

- Statement of the Problem: In practice, patients do not receive frequent testing

- Opportunity: Increase adherence to guideline-recommended testing; reduce cost

- Aim:
  - Identify patient and physician characteristics that predict a patient's likelihood of being tested in the next three months
  - Build an interactive software system designed to help clinicians and policymakers with identifying patients for CEA testing using the model.

- Data Source: de-identified Kaiser Permanente Southern California EHR database

User & Objectives: Physicians

Objectives:

- Address the requirements of colorectal cancer patients.
- Understand the likelihood of patient visits/CEA testing.
- Provide tailored treatment/resources (e.g., upstream intervention).

Interaction:

- Input necessary information using the interface.

Skill Level and Design Impact:

- Limited technical expertise.
- Interface must be user-friendly.
- Opportunity cost/Efficiency

Made by FREE-VECTORS.NET
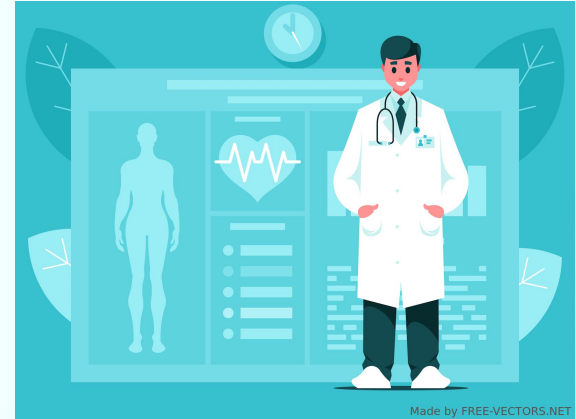
User & Objectives: Policy Makers (Governments, Health Providers, Insurers)

Objective:

- Utilize the interface for predicting colorectal cancer patients' clinical visits/CEA testing outcomes and implementing targeted interventions.
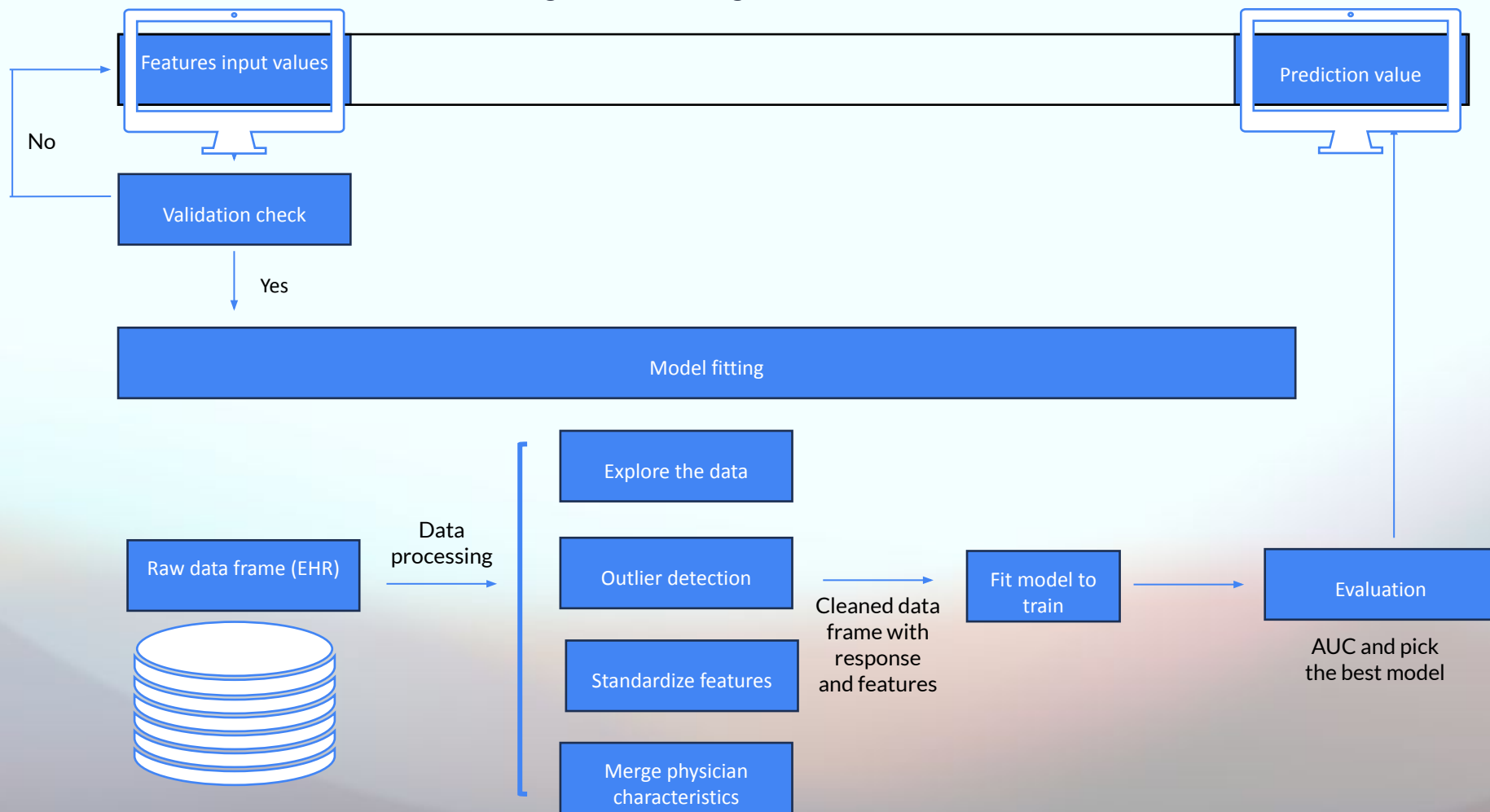
Interaction:

- Solely interact with the interface, inputting necessary information for predictions.

Skill Level and Design Impact:

- Limited technical expertise; interface should prioritize user-friendliness.
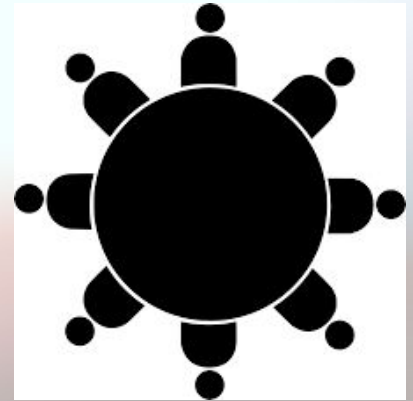
Component Diagram: Supporting Clinicians and Policy Makers

Demo: model_predict

Predictive Healthcare: Machine Learning for Optimized CEA Testing in Colorectal Cancer Patients

Predicting a patient's likelihood of being tested within the next three months

Enter the number of days from the last visit of the patient

Type a number

Enter the number of days from the patient's start of surveilliance

Type a number

Enter the number of days from the patient's first visit after surveilliance

Type a number

Enter the patient's Carcinoembryonic Antigen (CEA) value from the last visit

Folder Structure

```
├── LICENSE
├── README.md
├── data
│   ├── Final dataset prep_072521.csv
│   ├── default_output
│   │   ├── max_train.npy
│   │   ├── min_train.npy
│   │   └── model.pkl
│   ├── deid_cea_v2.csv
│   ├── deid_md_dep_v1.csv
│   ├── deid_md_edu_v1.csv
│   ├── deid_md_main_v1.csv
│   ├── deid_md_specialty_v1.csv
│   ├── modeltestfail.csv
│   ├── modeltestvalid.csv
│   └── modeltestvalid2.csv
├── Doc
│   ├── components.md
│   ├── technology_review
│   │   └── ML4CEA Technology review.pptx
│   └── userStories.md
├── examples
│   ├── example_model_predict.png
│   ├── model_predict.md
│   └── model_train.md
├── ml4cea
│   ├── __init__.py
│   ├── create_variable.py
│   ├── data_clean.py
│   ├── model.py
├── model_predict.py
├── requirements.txt
├── scripts
│   └── model_train.py
├── setup.py
└── tests
    ├── test_create_variable.py
    ├── test_data_clean.py
    └── test_model.py
```

Future direction

- Conduct an extensive model test to identify the most fitting and predictive model for

  our data (e.g., consider non-linear combinations of predictors)

- Host the website on a different server

- Publishing findings in academic journals

Questions?