ASTHMA PREDICTION TOOL

Group 7, HCDS
Free University of Berlin
Summer Semester 2024

PROJECT GOALS

- to develop a tool for physicians to use in the diagnosis and treatment of asthma patients

- the tool should assist with decision-making, be transparent and intelligible

TARGET AUDIENCE

doctors who diagnose patients
under **time constraints** and
do not have prior data science
knowledge

THE ASTHMA DISEASE DATASET

AUTHOR

Rabie El Kharoua

CONTENT

extensive health
information on
2,392 patients

SOURCE

KAGGLE

DATE

2024

MODEL TRAINING

26 FEATURES

demographics,
lifestyle and habits,
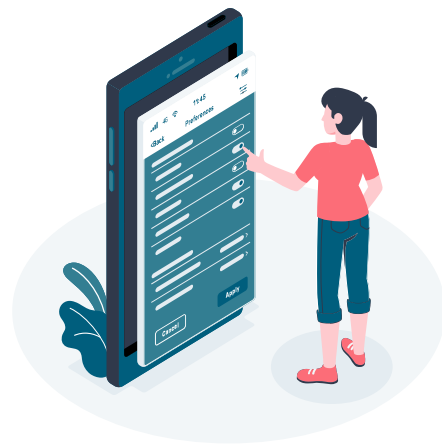medical history

PROTECTED VARIABLES

Gender, *Age*, *Ethnicity*,
FamilyHistory, and
EducationLevel

TARGET

Diagnosis,
imbalanced

SUPPORT VECTOR MACHINE

APP DEVELOPMENT
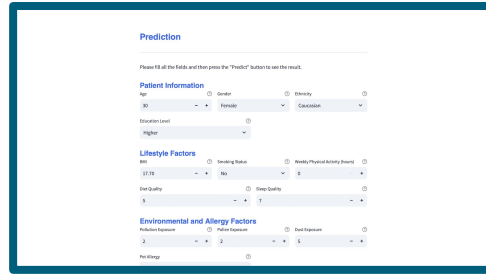
Creating a prototype

Peer feedback

Think-aloud study
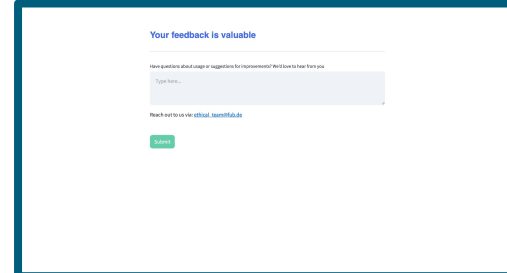
APP STRUCTURE

HOMEPAGE

PREDICTION

ABOUT DATASET

ABOUT MODEL

FEEDBACK

KEY DESIGN DECISIONS

INTUITIVE USER INTERFACE

- straightforward side menu

- a page overview to familiarize users with the content of each page

DISCLAIMER

Disclaimer:

This Asthma Prediction tool is an aid tool and should not be fully relied upon for diagnosis. It is intended to assist doctors by providing additional insights. Please consider the points below before starting to use the tool:

1. When in doubt about the fairness of the results, alter the values of protected variables (e.g., gender, ethnicity) to see if the result changes significantly.
2. Review the distribution of the training dataset used before using the tool for reliability check (e.g., ensure the ethnicity group is included or check if the patient age is too old/young compared to the model's training dataset).
3. Needed test results: Ensure that the patient has undergone the necessary tests (e.g., Lung Function FEV1 and FVC) as they are required for the model's prediction.
4. For checking if the model aligns with medical knowledge, please refer to the global explanation section in the 'About Model' page to understand which variables were most significant for prediction.

- displayed at the beginning of the homepage to ensure that users see it before using the tool

- highlights all the important points users need to consider before using the tool
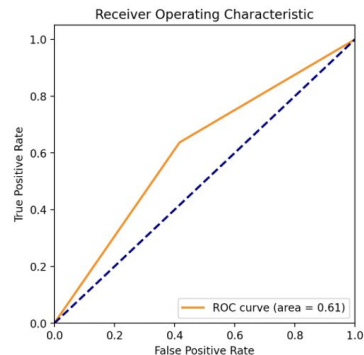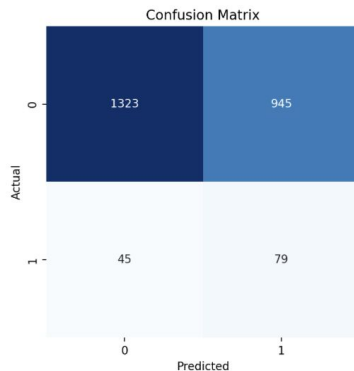
DETAILED EXPLANATIONS

the tool provides very detailed explanations of various data science concepts

Model Performance Evaluation

Accuracy: 0.59

	No Asthma	Asthma
precision	0.9671	0.0771
recall	0.5833	0.6371
f1-score	0.7277	0.1376

Confusion Matrix

Receiver Operating Characteristic — ROC curve (area = 0.61)

Interpretation

- **Accuracy**: Measures how often the model is correct. High accuracy indicates the model performs well.
- **Precision**: Indicates how many of the positive predictions are actually correct.
- **Recall**: Indicates how many of the actual positives the model correctly identified.
- **F1-Score**: Harmonic mean of precision and recall. High F1-score indicates a balance between precision and recall.
- **Confusion Matrix**: Shows the counts of true positive, true negative, false positive, and false negative predictions.
- **ROC Curve**: Graphical representation of the true positive rate vs. the false positive rate.
- **AUC (Area Under the Curve)**: Measures the overall ability of the model to discriminate between positive and negative classes. Higher AUC indicates better performance.
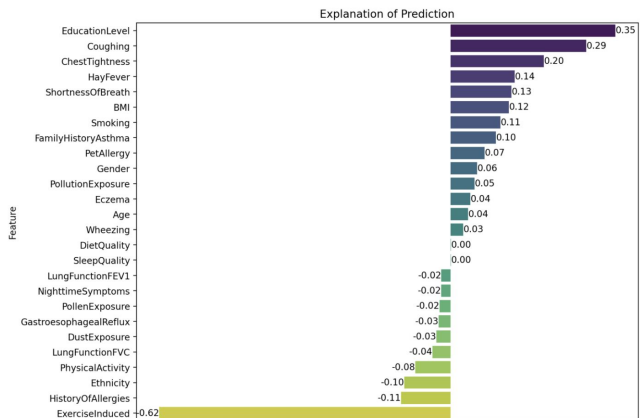
Dataset Similarity Check
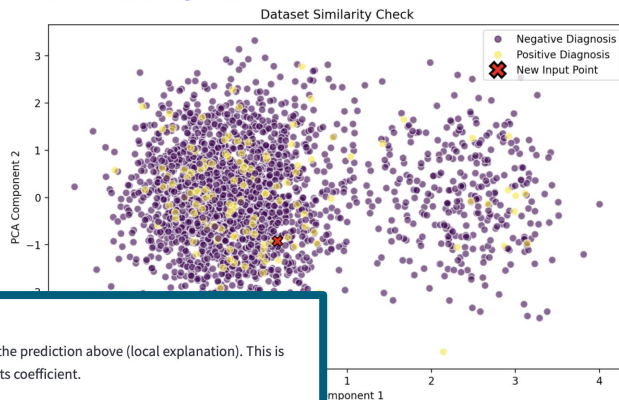
Dataset Similarity Check

Explanation of Prediction

The plot below shows the contribution of each feature to the prediction above (local explanation). This is calculated as the product of the scaled feature value and its coefficient.

Explanation of Prediction

Features with higher contributions are more influential in the prediction. This helps to understand which features are pushing the prediction towards a positive or negative diagnosis.

RELIABILITY OF PREDICTIONS

the tool offers multiple methods to help doctors assess the reliability of the prediction

FAIRNESS EVALUATION

the application includes a fairness evaluation section to ensure the model treats different demographic groups equitably
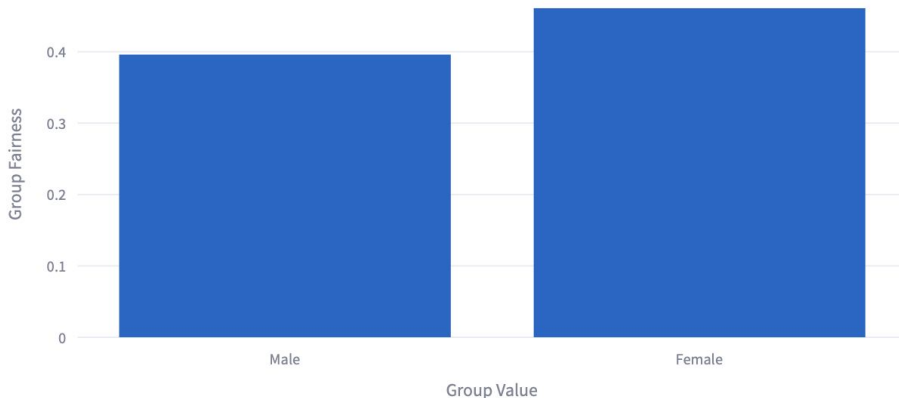
Select fairness metrics:

Group Fairness

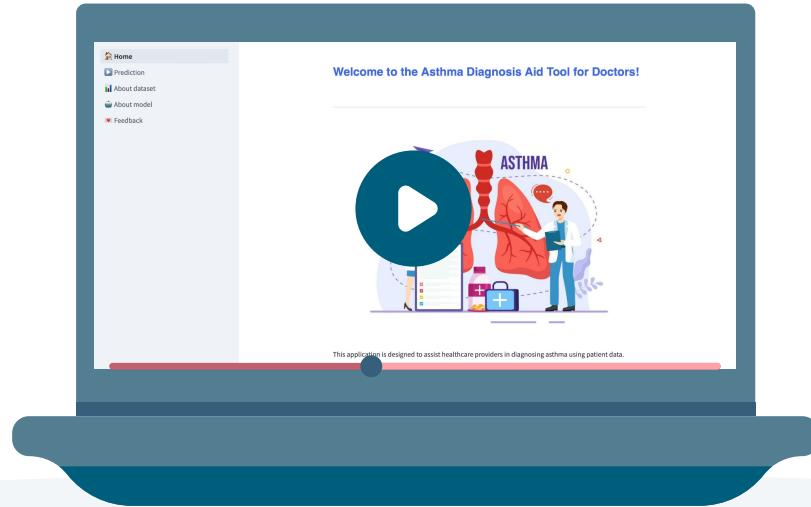Select a protected attribute for analysis:

Gender

Select whether No Asthma or Asthma should be considered as positive class.

Asthma

Group Fairness for protected variable Gender

LIVE DEMO OF THE INTERFACE

CHALLENGES & REFLECTIONS

Imperfection of the dataset

Large amount of information

Complexity of explanations

brainstorming

peer feedback

UI test

THANKS

Do you have any questions?