Aditya Panchal   Mariana Steffens   Navya Reddy Tiyyagura   Se Yeon Kim

Human Centered Data Science SS24

Membership: Group 3

Project Descriptions

Github Repository: https://github.com/marianast97/Ethics_Project.git

1. **Project goal:**
   Create a comprehensive explanation interface using the techniques learned in class to help our target audience make informed decisions on maternal health.

2. **Intended target audience:**
   Our target audience are the medical staff (mainly nurses) in a maternity hospital, who work in the triage process to define the urgency of each patient.

3. **The Dataset and documentation:**

Chosen Dataset: Maternal Health Risk

Ahmed, Marzia. (2023). Maternal Health Risk. UCI Machine Learning Repository. **https://doi.org/10.24432/C5DP5D**.

Related Studies:

M. Ahmed and M. A. Kashem, "IoT Based Risk Level Prediction Model For Maternal Health Care In The Context Of Bangladesh", 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2020, pp. 1-6, doi: 10.1109/STI50764.2020.9350320.

Summary of the dataset's content and its key features

Data was collected from five hospitals and one maternity clinic in Dhaka, Bangladesh. Patient health data was collected using wearable sensor devices, and risk levels were classified with the help of medical experts and literature review.

Variable Name	Role	Type	Description	Units	Missing Values
Age	Feature	Integer	Age in years when a woman is pregnant.		no
SystolicBP	Feature	Integer	Upper value of Blood Pressure in mmHg, a significant attribute during pregnancy.		no
DiastolicBP	Feature	Integer	Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.		no
BS	Feature	Integer	Blood glucose levels, in terms of molar concentration	mmol/L	no
BodyTemp	Feature	Integer		F	no
HeartRate	Feature	Integer	A normal resting heart rate	bpm	no
RiskLevel	Target	Categorical	Predicted Risk Intensity Level during pregnancy considering the previous attributes.		no

4. **Key design decisions we made:**

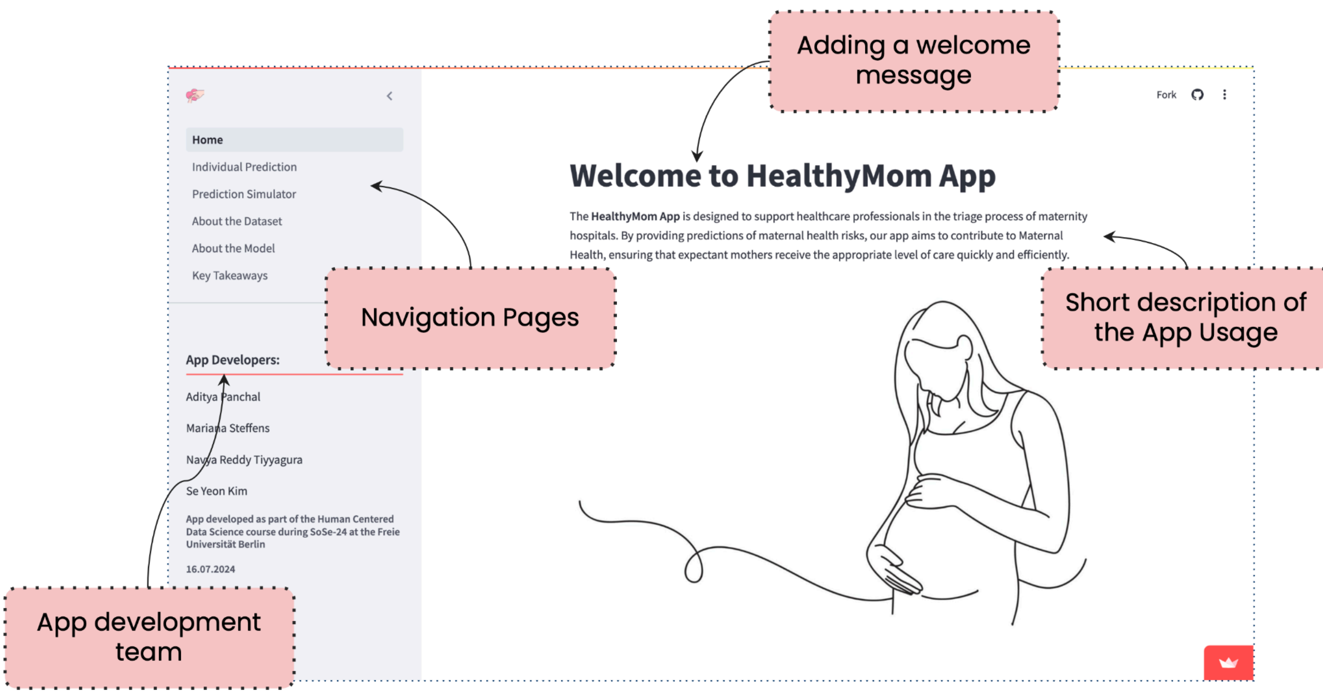
To guide us through the design process of the interface, we formulated some important questions that the user should be able to answer when using our application. The questions are listed below:

- How each feature contributed to the prediction of the mother_id 31?
- How would the prediction change if we change the feature inputs for the mother_id 31?

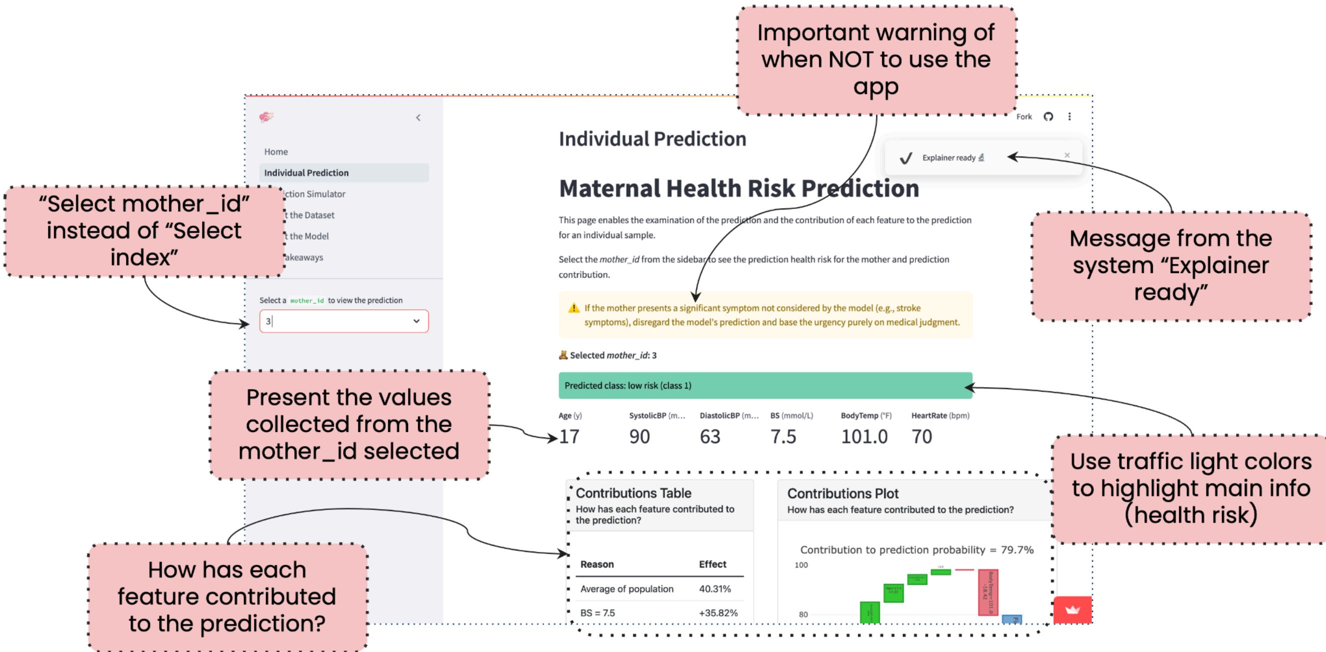
- How was the data collected?
- What are the main features of the classification model?
- What is the uncertainty in the prediction?

Based on those questions, we decided to divide our app in different pages:

- **Home:** Welcomes users and describes the app's purpose.
- **Individual Prediction:** Allows medical staff to select a mother ID and view the health risk prediction (high risk, medium risk, low risk).
- **Prediction Simulator:** Enables users to modify input feature values and observe how predictions change.
- **About the Dataset:** Provides comprehensive information about the dataset, including data collection, labeling, and attribute distribution.
- **About the Model:** Offers details about the model used and its performance, intended for both medical staff and data scientists.
- **Key Takeaways:** Summarizes our reflections on appropriate and inappropriate use cases for the model.



We've decided to add important warnings on the main pages "Individual Prediction" & "Prediction Simulator" to create awareness specifically on when NOT to use the model. We've decided to use a lot of traffic light color coding to highlight the prediction risk: red for high risk, yellow for medium risk and green for low risk. We've added messages from the system to provide some feedback to the user when, for example, the explainer was loaded or the training was completed.



To make it easier for the user to understand the prediction and the prediction contribution, we decided to add some text which explains to the user how the table/visualization should be read.

The table shows the contributions of each feature to the prediction for the selected sample. (Higher the 'Final Predction', lower the predicted risk level)

And the plot shows how the model makes a prediction for a sample i.e. how each feature contributes to the prediction.

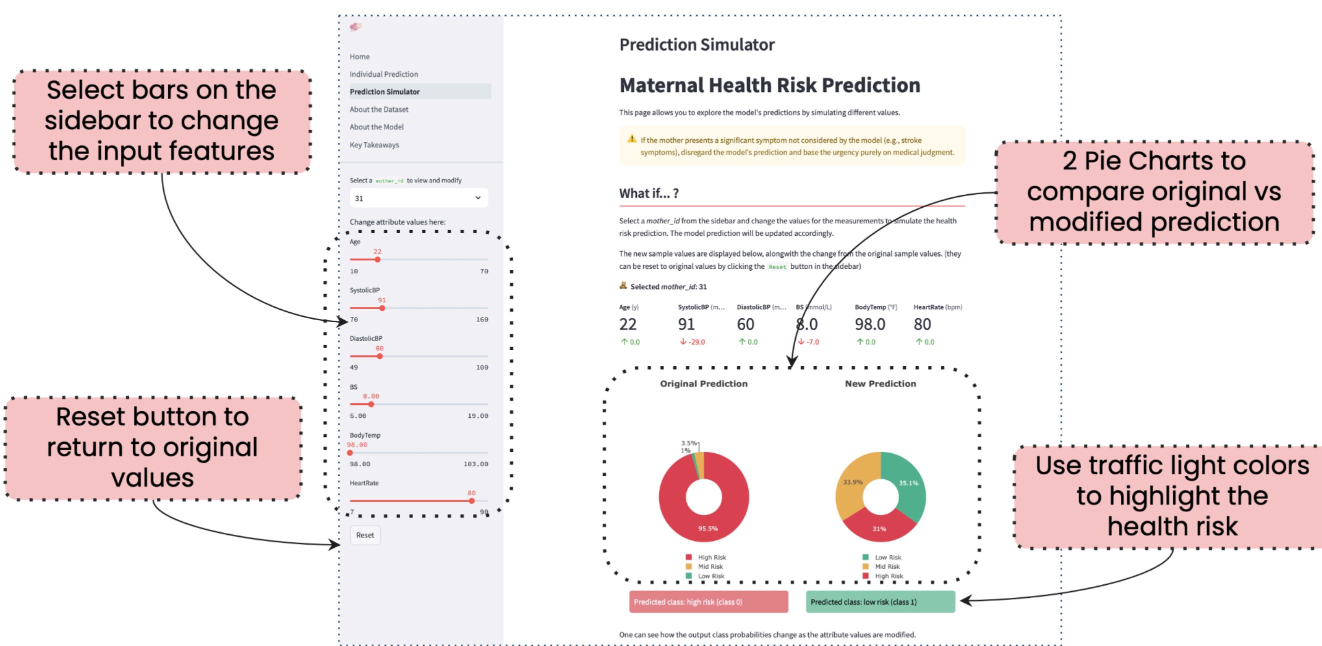
We start from the left that is the baseline a.k.a the average prediction of our model. We then add all the contributions given by the SHAP values one by one and until we end up with the final prediction the right.

A higher value of the 'Prediction %' indicates a final prediction of low risk, while a lower value indicates a final prediction of high risk.

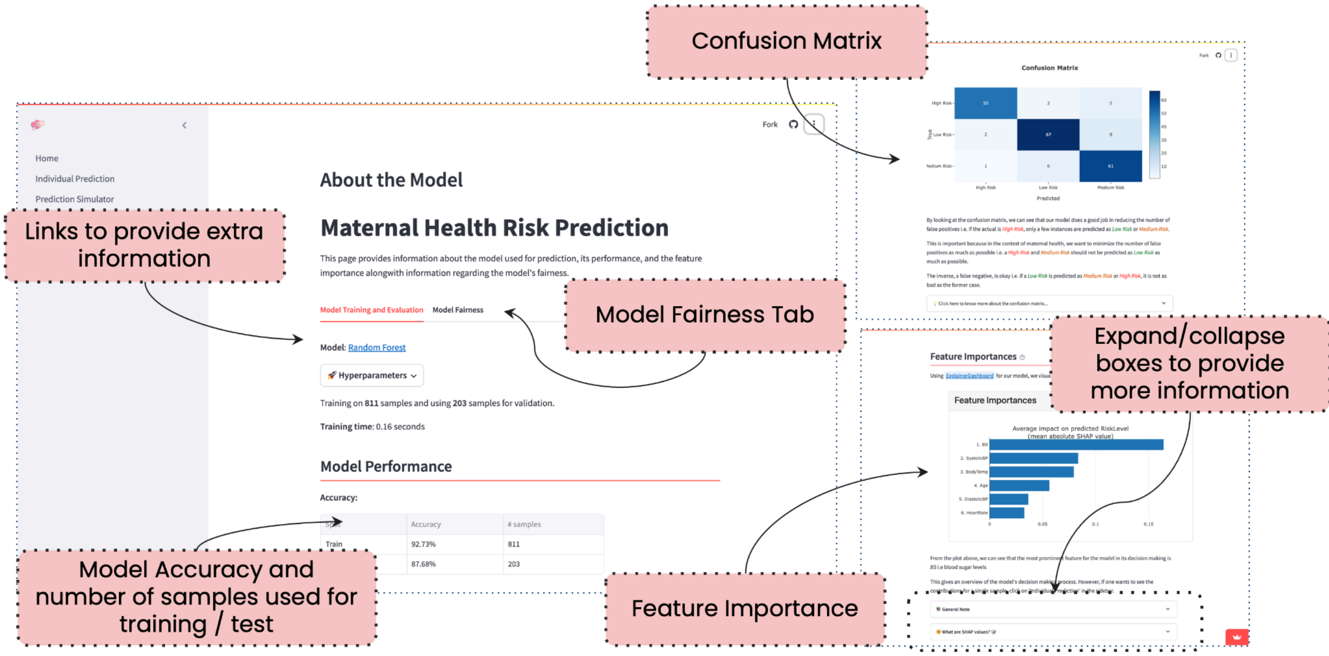
⚠️ **Note:** Before making any decision based on the model's final prediction, please also have a good look at the `Confusion Matrix` under the 'About the Model' page. It gives necessary information about the model's performance, namely *false positives* and *false negatives*.

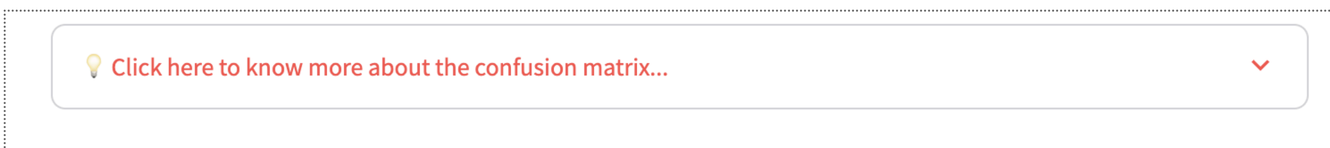
ℹ️ One of the fundamental properties of Shapley values is that their sum yields the difference of actual and average prediction. In our case, it means that SHAP values of all the input features will always sum up to the difference between baseline (expected) model output and the current model output for the prediction being explained. Also, the baseline remains the same for each point as it is simply the mean of all the predictions of our model

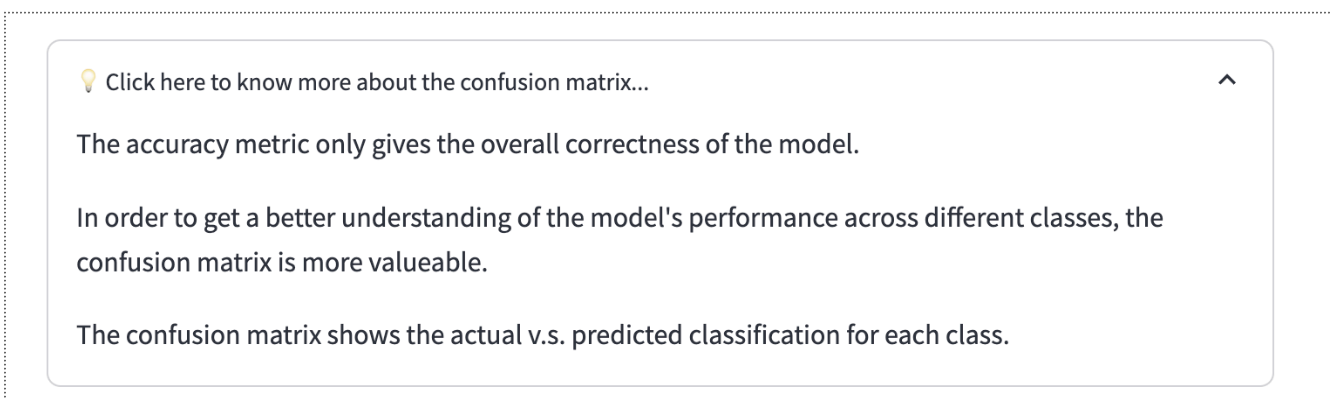
To explore the prediction based on different input values, we've added the page "Prediction Simulator". There, the user is able to change the input features on the sidebar and on the right side, the user can compare the original prediction with the new one.



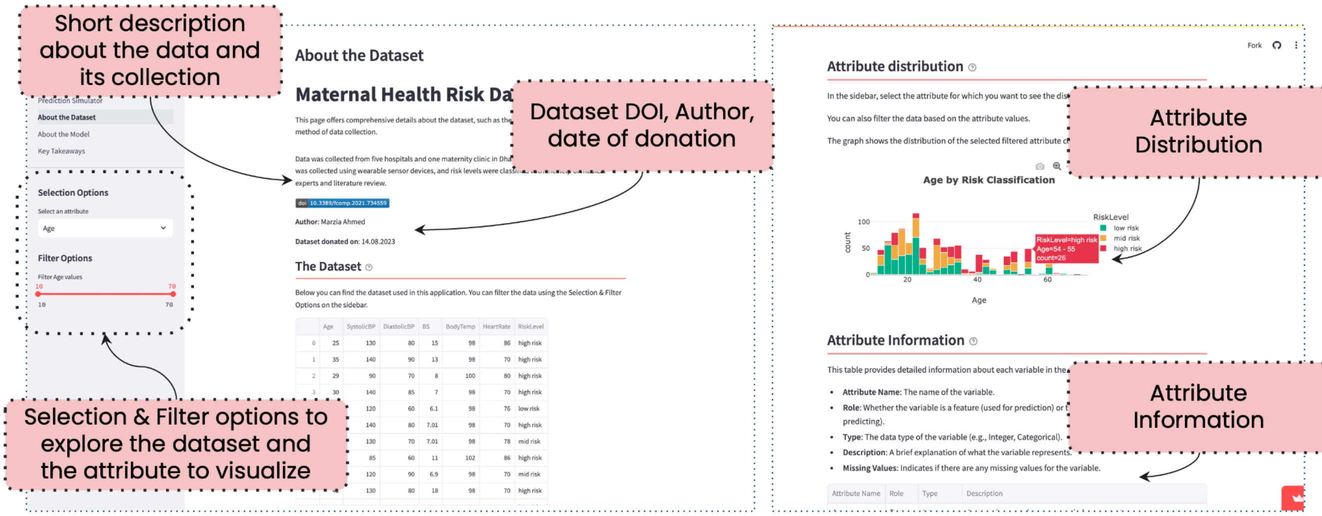
To provide more details without overwhelming the user with too much information, we used boxes that collapse and expand when the user would like to see more information.



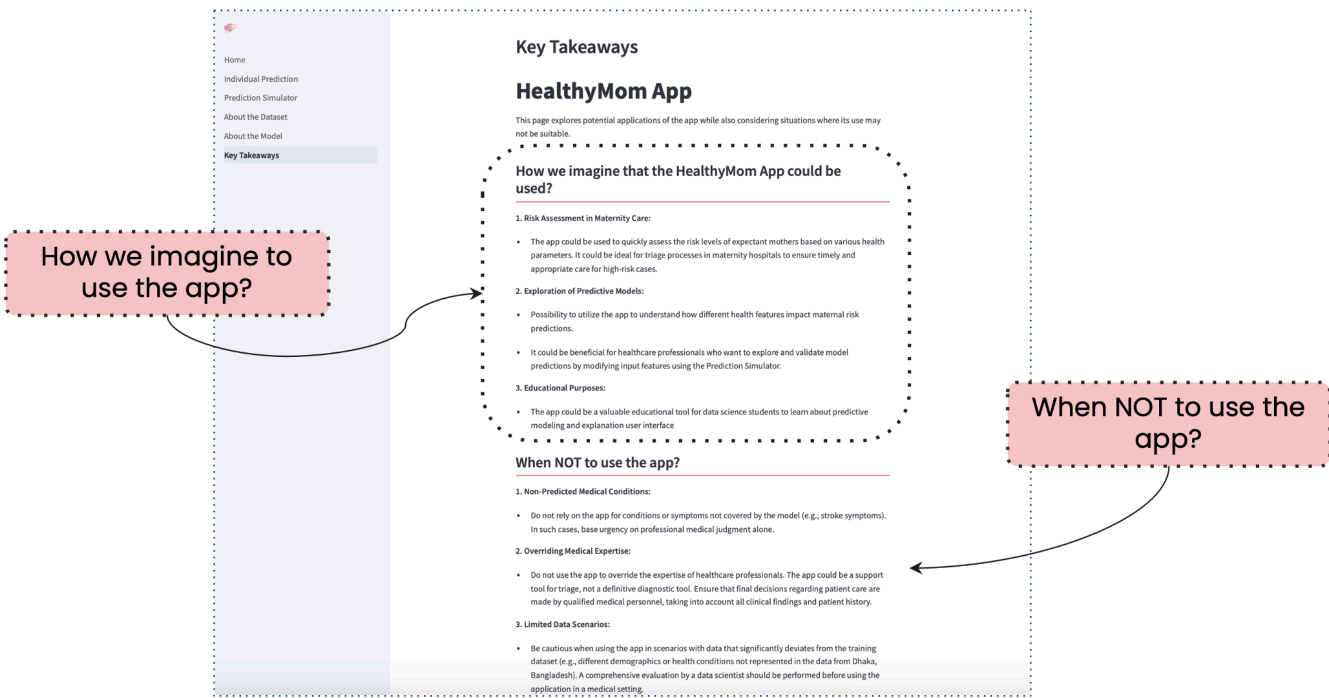




To allow the user to explore the dataset, we decided to add the complete dataset to the About the Dataset page, including the visualization of the attribute distribution. On the left sidebar, the user can select the attribute for which he or she would like to see the distribution. Like that, we avoid adding too many visualizations at once.



The last page of the app provides a reflection of how we imagine that the app should be used and in which situations that app should not be used.



5. Problems we faced:

Our main problem was with the fairness metrics. When we selected the dataset, we planned to use the fairness attributes on the age group, but afterwards we realized that it made sense to discriminate on the age group, since this is an important risk factor in pregnancy. The dataset did not have any other attribute that made sense for assessing fairness. In the end, we decided to use the age group fairness metrics to show that the model was indeed discriminating on the basis of age.

Another issue that we faced was regarding the interface provided by the explanation dashboard. In our opinion, the dashboard contains too much information that could overwhelm the user, especially if the user does not have a background in data science, for example. To address this issue, we decided to select only the modules and the visualization that in our opinion, would be relevant for our stakeholder.

6. **Reflection of the development process:**

This project allowed us to design an explanation user interface that meets user needs for explainability while ensuring usability and effectiveness. We also learned to work as a team, planning activities and dividing the workload. Coordinating the group and tracking task completion were essential for achieving the final product. It's important to note that this app was developed solely for educational purposes and did not involve a real stakeholder or use case. We made several assumptions about the triage process in a hospital, based on our imagination. Although the application is not usable in reality, its development provided valuable insights into ethical data science practices and human-centered design.