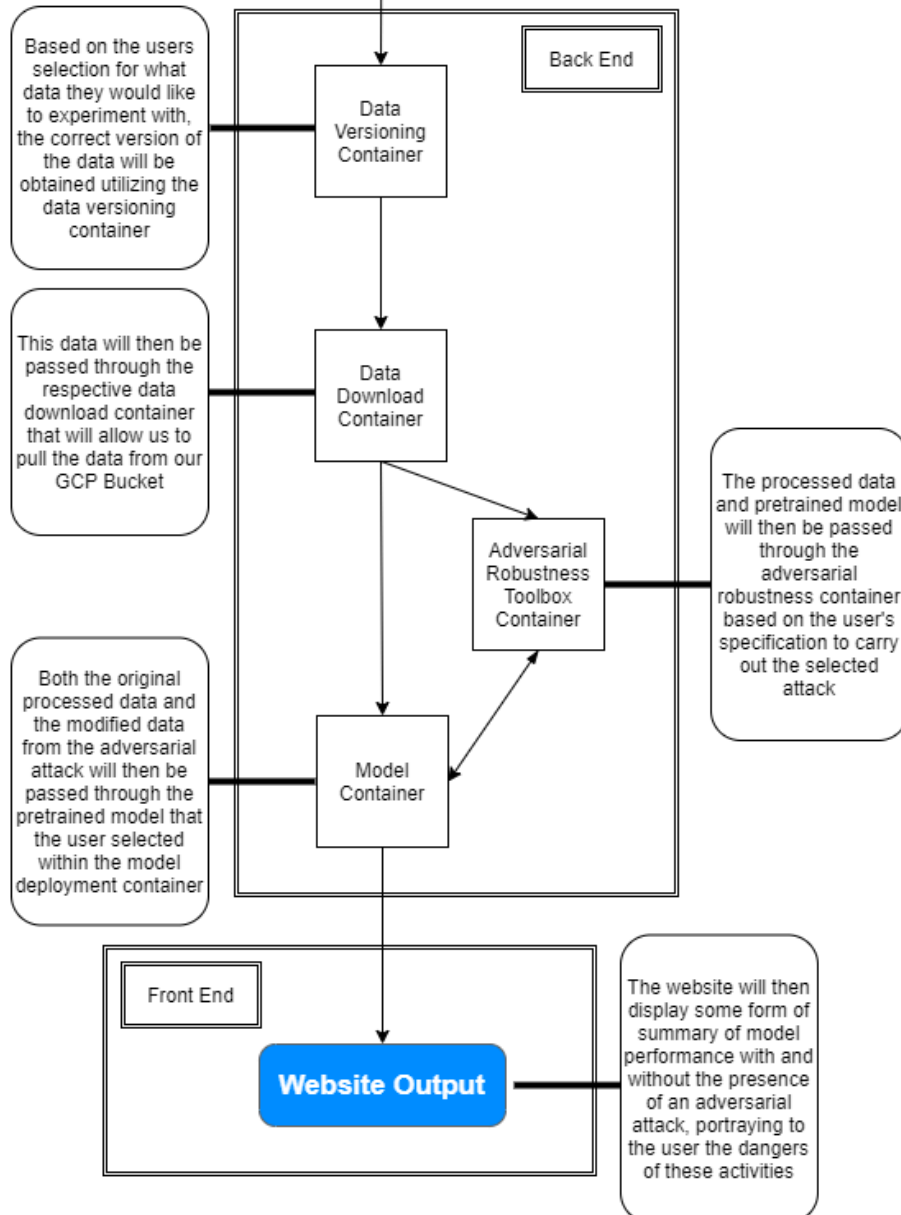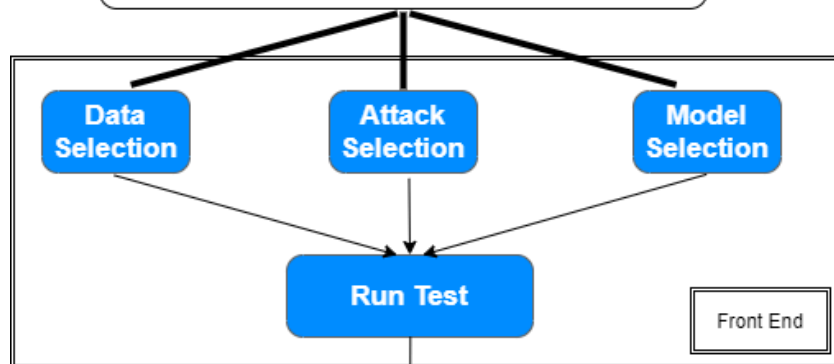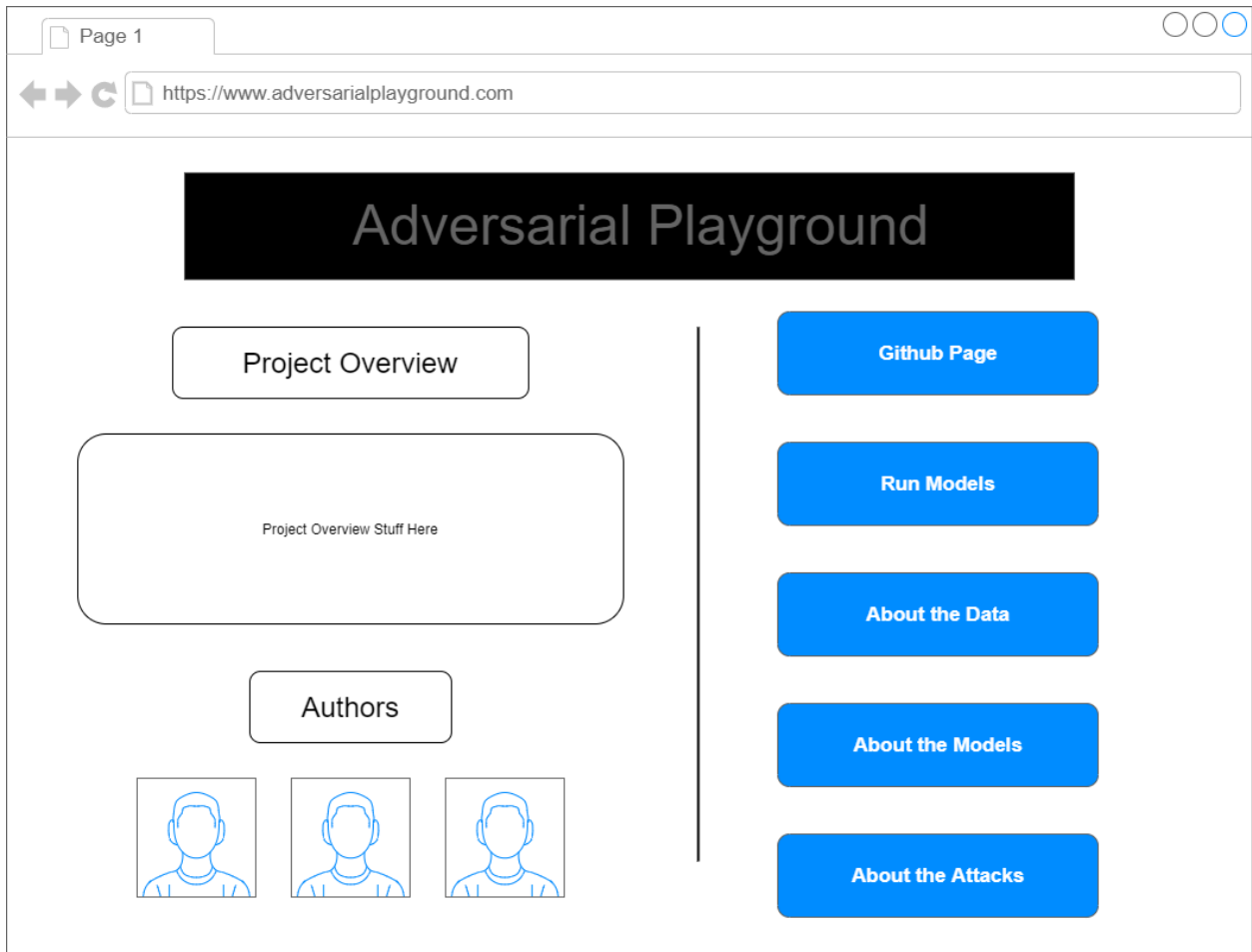On the webpage, the user will be able to specify the type of data that they want to use, the attack which the want to employ, and the model that they want to test the attack on (this will be automatically selected based off of the dataset selection)

Data Selection

Attack Selection

Model Selection

Run Test

Front End

Back End

Based on the users selection for what data they would like to experiment with, the correct version of the data will be obtained utilizing the data versioning container

Data Versioning Container

This data will then be passed through the respective data download container that will allow us to pull the data from our GCP Bucket

Data Download Container

Adversarial Robustness Toolbox Container

The processed data and pretrained model will then be passed through the adversarial robustness container based on the user's specification to carry out the selected attack

Both the original processed data and the modified data from the adversarial attack will then be passed through the pretrained model that the user selected within the model deployment container

Model Container

Front End

Website Output

The website will then display some form of summary of model performance with and without the presence of an adversarial attack, portraying to the user the dangers of these activities

https://www.adversarialplayground.com

Adversarial Playground

Project Overview

Project Overview Stuff Here

Authors

Github Page

Run Models

About the Data

About the Models

About the Attacks

https://www.adversarialplayground.com/run-models

Run Models

Return to Homepage

Experiment Selection

Dataset Selection	Model Selection	Attack Selection
HAM10000	ResNet-101	Attack 1
GTSRB	YOLOv5	Attack 2

Run Test

Performance Results

This will be uploaded with some form of performance comparison and possibly an example image of what the difference between running the models on the test set without the attacks taking place and then with the attacks taking place looks like

Data Overview

Return to Homepage

Traffic Sign Dataset

The Traffic Sign Dataset has over 500 images that have corresponding bounding box specifications that can be used to fine-tune YOLO models. The data set has four traffic sign classes: prohibitory, danger, mandotory, and other.

Link: https://www.kaggle.com/datasets/valentynsichkar/traffic-signs-dataset-in-yolo-format/data

HAM10000

The Humans Against Machines 10000 (HAM10000) dataset is a single-class, single-image classification dataset which contains information on seven different kinds of skin cancer. It contains over 10000 training images and roughly 2000 testing images, with each image being tagged with the specific type of skin cancer or lesion that is present within the image.

Link: marmal88/skin_cancer · Datasets at Hugging Face

https://www.adversarialplayground.com/models

Model Overview

Return to Homepage

ResNet-101

YOLOv5

ResNet-101 is a model that was pre-trained on ImageNet at a resolution of 224x224. It is a CNN that utilized residual learning and skip connections for image classification, which has been fine tuned for our project purpose in such a way so that the final output of the model contains the predictions of 7 different classes, being the classes within the HAM10000 dataset.

YOLOv5 is an iteration within the Ultralytics YOLO series, designed to perform real-time object detection. It specializes in object detection tasks and has been fine tuned to fit the needs of the Traffic Sign dataset.

Attack Overview

Return to Homepage

Attack 1	Attack 2
Information about Attack 1	Information about Attack 2