



Auto Labeler

Matthew Beech, Isaac Bland, Trey Hedrick, Daniel Holman, Briley Johnson. Gage Smith



Description of Tool

The NavSEA Auto Labeler is a web interface tool that helps Data Science Researchers tackle the problem of not having enough labeled data for Supervised Learning AI/ML. The tool allows Researchers to upload unlabeled datasets that Labelers can then label. Both Researchers and Labelers are able to create accounts customized for their role in the tool (**Figure 1**). In addition, the tool includes a built-in Machine Learning System that allows Researchers to automate their research with the uploaded data.

Role of Labeler

- Help Researchers label the datasets they upload
- View random pieces of data and provide labels for them (**Figure 2**) in one of two ways:
 1. Typing in the blank
 2. Choosing from a list of options

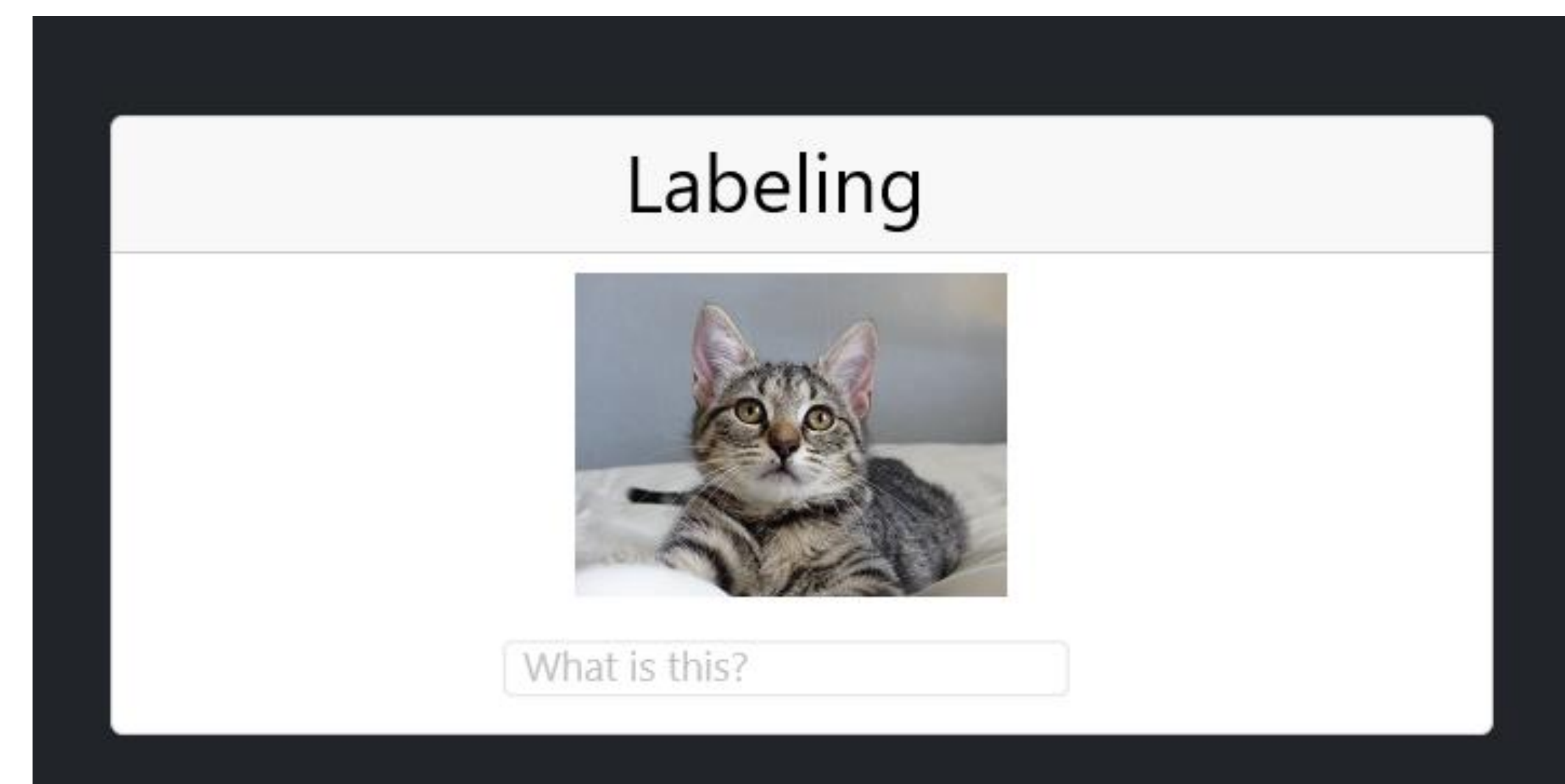


Figure 2. Data Labeling Page (Labeler). Here, the Labeler is shown an image of a cat and is prompted to type in what it is.

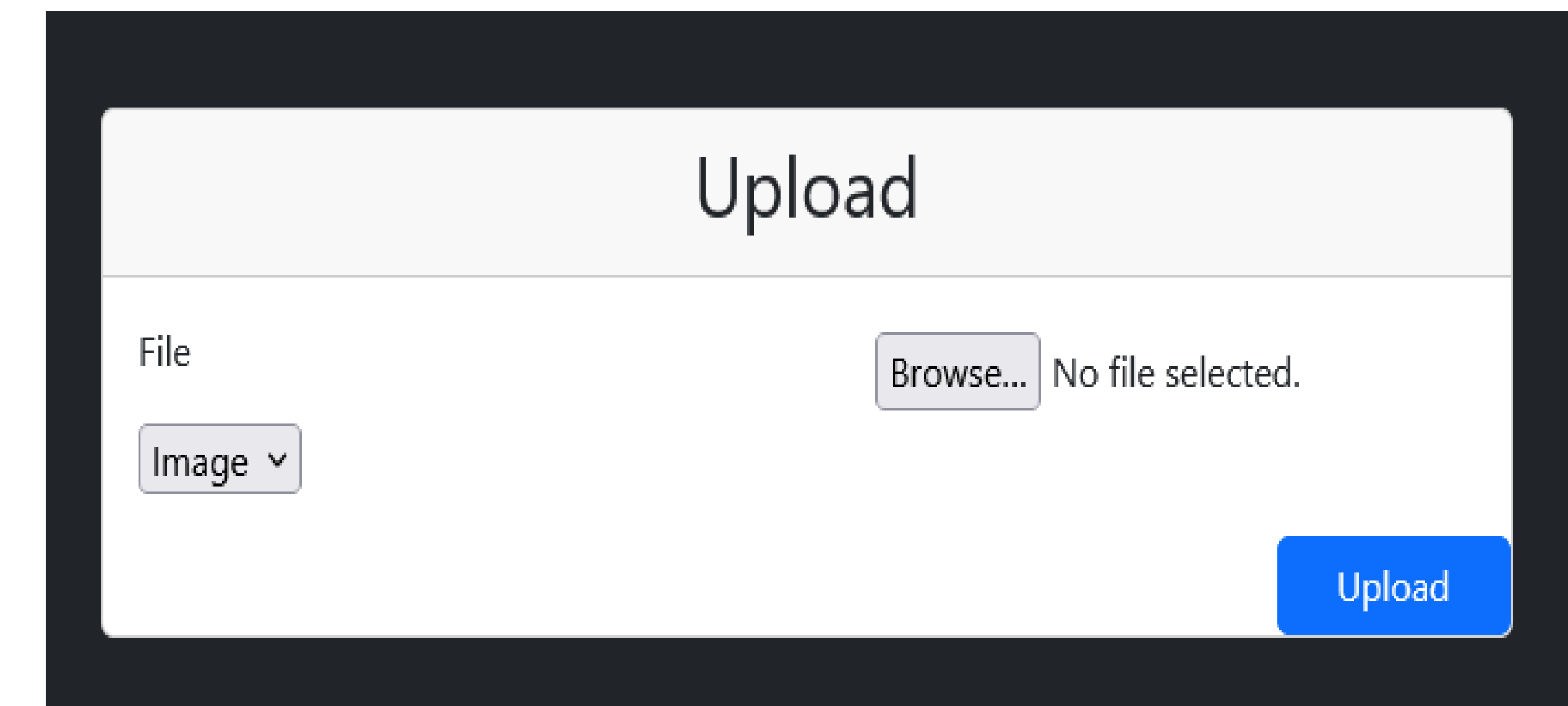


Figure 3. Dataset Upload Page (Researcher). The Researcher is able to browse files on his/her computer and select a dataset to upload (to be labeled by Labelers).

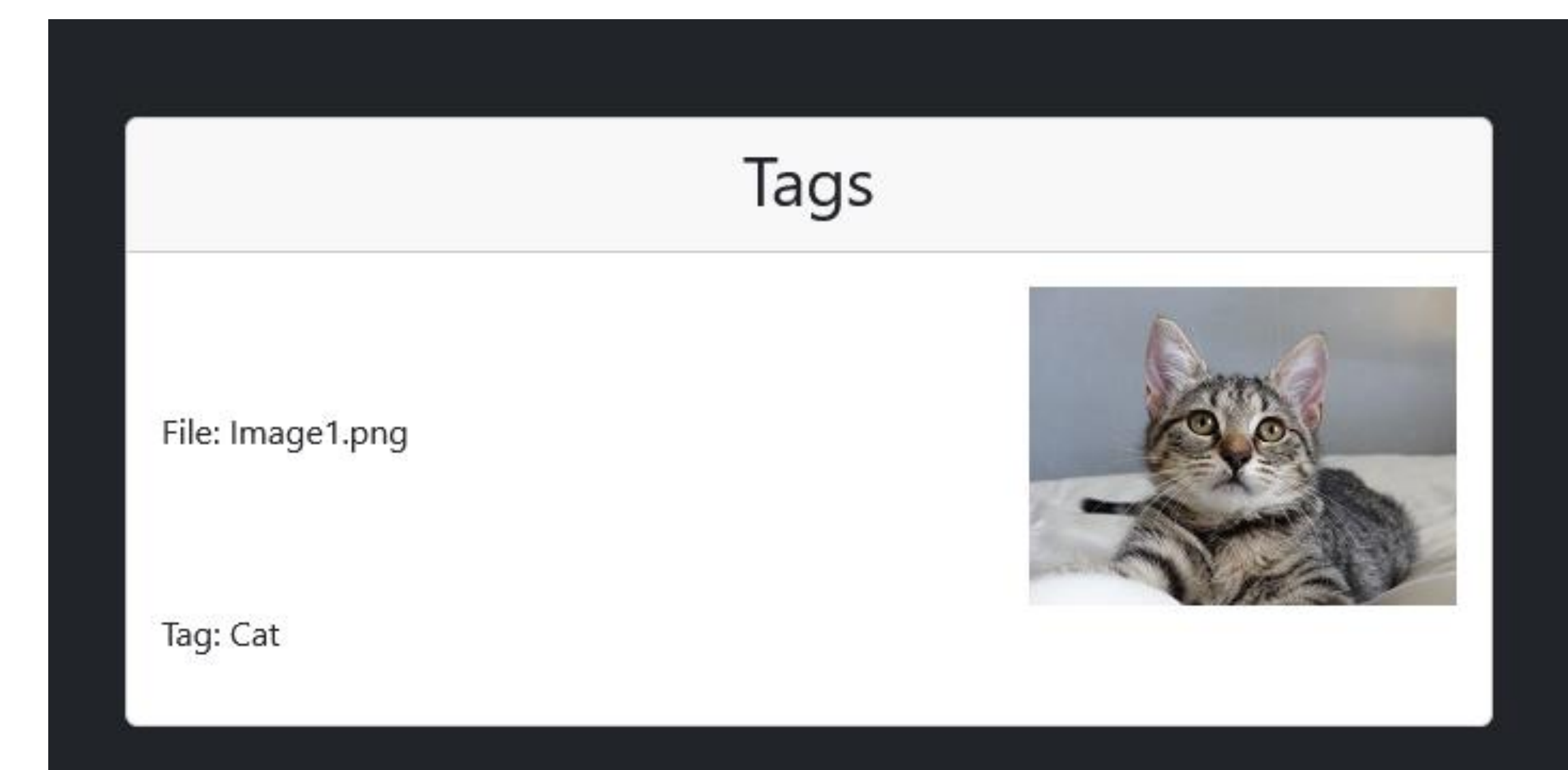


Figure 4. Data Tag Viewing Page (Researcher). Here, the Researcher is shown the single label that Labelers have selected for this image ("Cat").

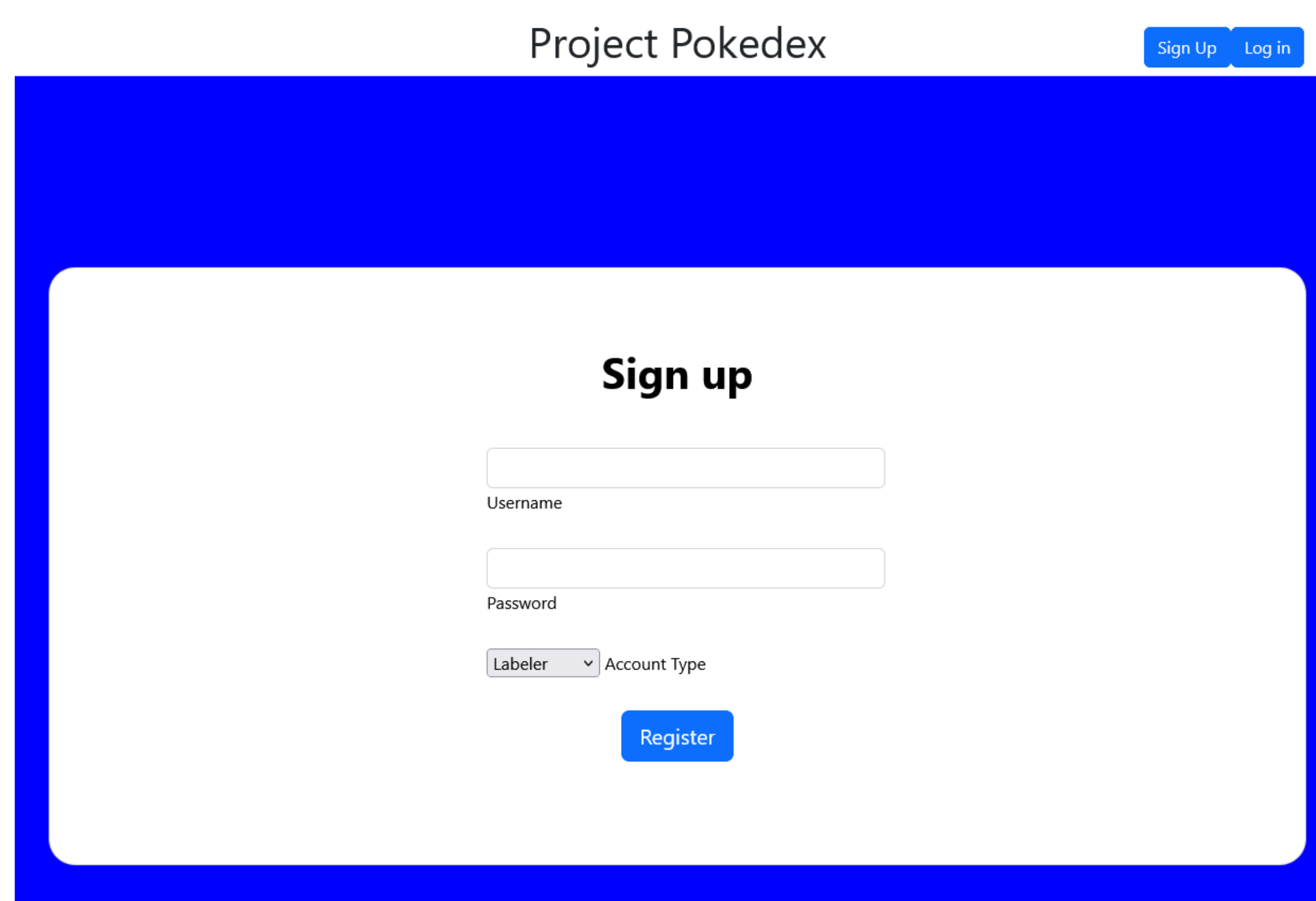


Figure 1. Account Creation Page. Users create a username and password and select an account type (Labeler or Researcher).

Role of Researcher

- Upload datasets containing image, audio, video, or text data (**Figure 3**)
- Re-download the datasets with all labels assigned
- View and track all labels (tags) assigned to pieces of data in each dataset (**Figure 4**)

Features of Built-in Machine Learning System

- Imports uploaded datasets then trains on their labeled data to assign labels to their unlabeled data
- Chooses the highest-scoring model (Decision Tree or Logistic Regression) to assign the new labels
- Uploads the new labels to the dataset, which can then be used as future training data

