Katz
Katz School
of Science and Health

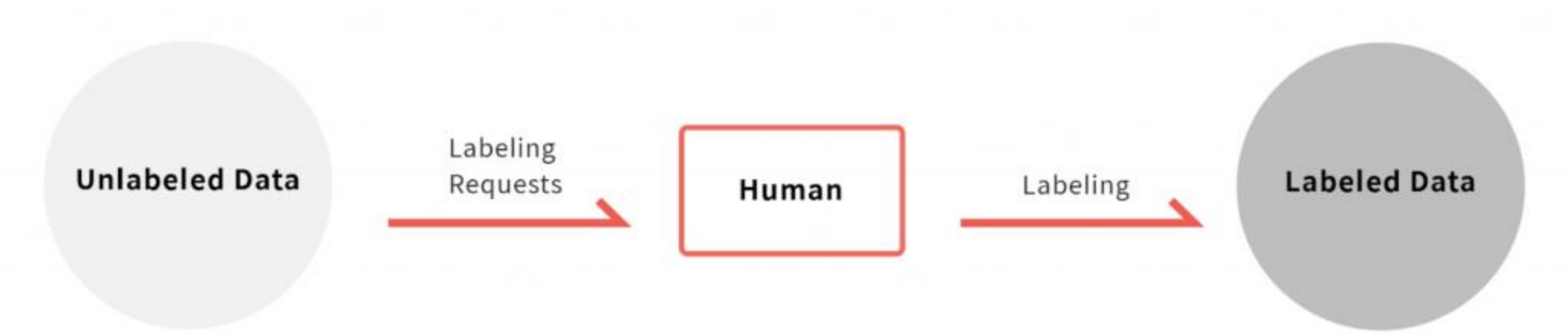
Yeshiva University - Katz School of Science

Data Analytics and Visualization MS

date: May 2022
from: Jesus Olivera, Independent Study

subject: YU Hate Speech Project Final Presentation

Designing and Implementing a V1 Data Labeling Tool Backend for Building a Training Dataset for ML/AI Hate Speech Classification



Introduction

In recent years, American Jews have faced increased threats of violence and harassment off and online. According to ADL (Anti-Hate Crime Organization) Antisemitic incidents in 2019 and 2020 were, respectively, the highest and third-highest years on record for cases of harassment, vandalism, and assault against Jews in the United States since the organization started tracking these incidents back in 1979.

In the past few years , tech platforms have focused resources on trying to combat online hate speech. This project is designed to help advance the efforts around detecting and preventing Antisemitic online hate speech by providing Yeshiva University's Research Department a framework for data pre-processing and data flows for the development of Machine Learning and AI capabilities aimed at detecting and preventing online Antisemitic hate speech.

nti-Semitic; also spe
one or something that is
to or prejudiced against
Semitic beliefs were well-
anti-Semitism
Semitism is hostility
Jewish people. The ext
Front promoted anti-se

Katz
Katz School
of Science and Health

Problem Statement & Objective

Antisemitism is on the rise. New technologies and applications are required to combat the negative effects of online hate speech towards the Jewish Community; For this reason, the project will be focused on a complex generic use case for an MVP aimed at designing, building, and deploying the backend of a V1 data labeling pipeline that builds robust labeled datasets that can be used to train Machine Learning and AI algorithms to detect online antisemitic hate speech.

Katz
Katz School
of Science and Health

Importance & Broader Impact

Developing techniques and methods for combating online hate speech can provide tech companies with a framework around monitoring content in their platforms. Overtime, these initiatives can enhance regulatory compliance efforts and ultimately provide users a safe space for their regular interactions within online platforms.

Mental health has declined over the years and some of the variables that have impacted the overall uptake in suicides, depression and other mental health problems can be attributed to the rise in technology and in particular social media.

A future state for this project can include providing access to external organizations, via a data marketplace or a deployment space, to high-quality antisemitic labeled datasets for their models and/or access to trained models focused on detecting antisemitism online.

This project has the potential of contributing to the overall benefit of society by mitigating some of the negative effects of online hate speech towards Jews and the broader communities.

Katz

Katz School
of Science and Health

YU Hate Speech Backend Engineering Focused MVP Design

Sprint 1

High-Level Project Development Phase 1:

Area of Focus: Backend V1

Timeframe: January 2022 to May 2022

Owner: Jesus Olivera, Lead Engineer

- ✓ Literature and Existing Components Review
- ✓ Gap Analysis
- ✓ Tools Selection
- ✓ Process Logic Development
- ✓ Architecture Design
- ✓ Pipelines Design & Deployment Instructions

Sprint 2

High-Level Project Development Phase 2:

Area of Focus: ML / AI

Timeframe: May 2022 to TBD

Owner: TDB

- Model Design
- Model Development
- Model Deployment
- Model Inputs & Outputs Process Logic Design, Development & Implementation
- Model Setup

Sprint 3

High-Level Project Development Phase 3:

Area of Focus: User Interface

Timeframe: TBD

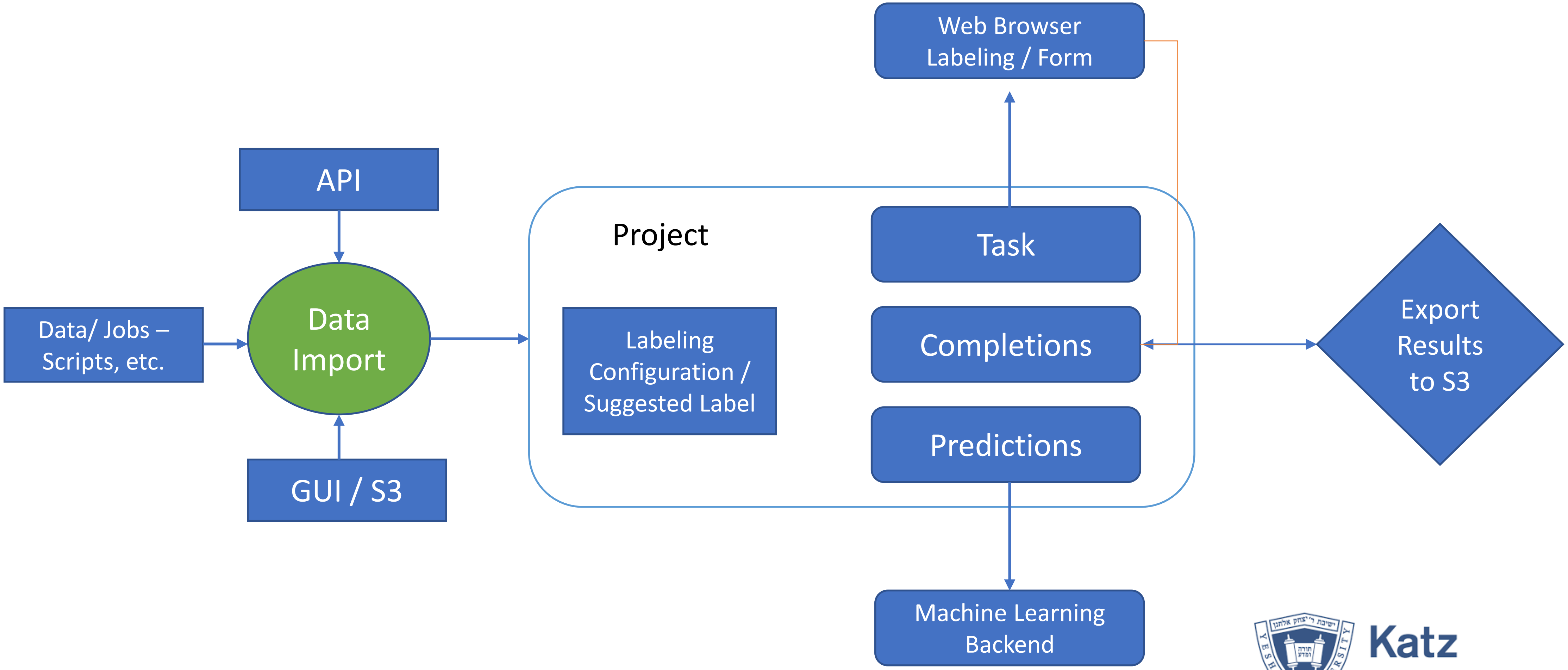
Owner: TDB

- Design Thinking: Understanding users and their values
- Competition Research
- Interface Sketch/ Wireframe
- Interface Design
- Design Implementation
- Implementation Evaluation

Katz

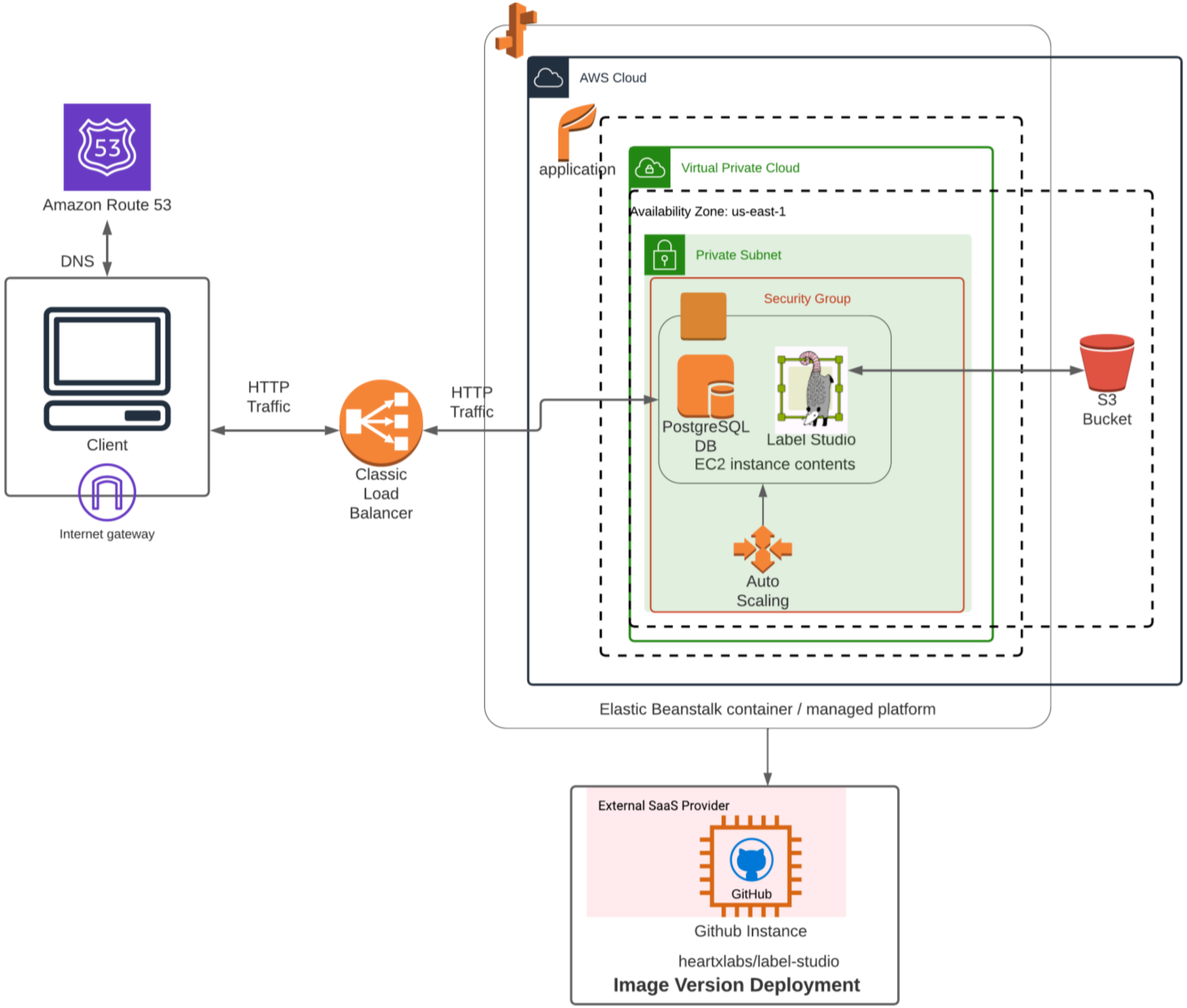
Katz School
of Science and Health

Application High-Level Logic



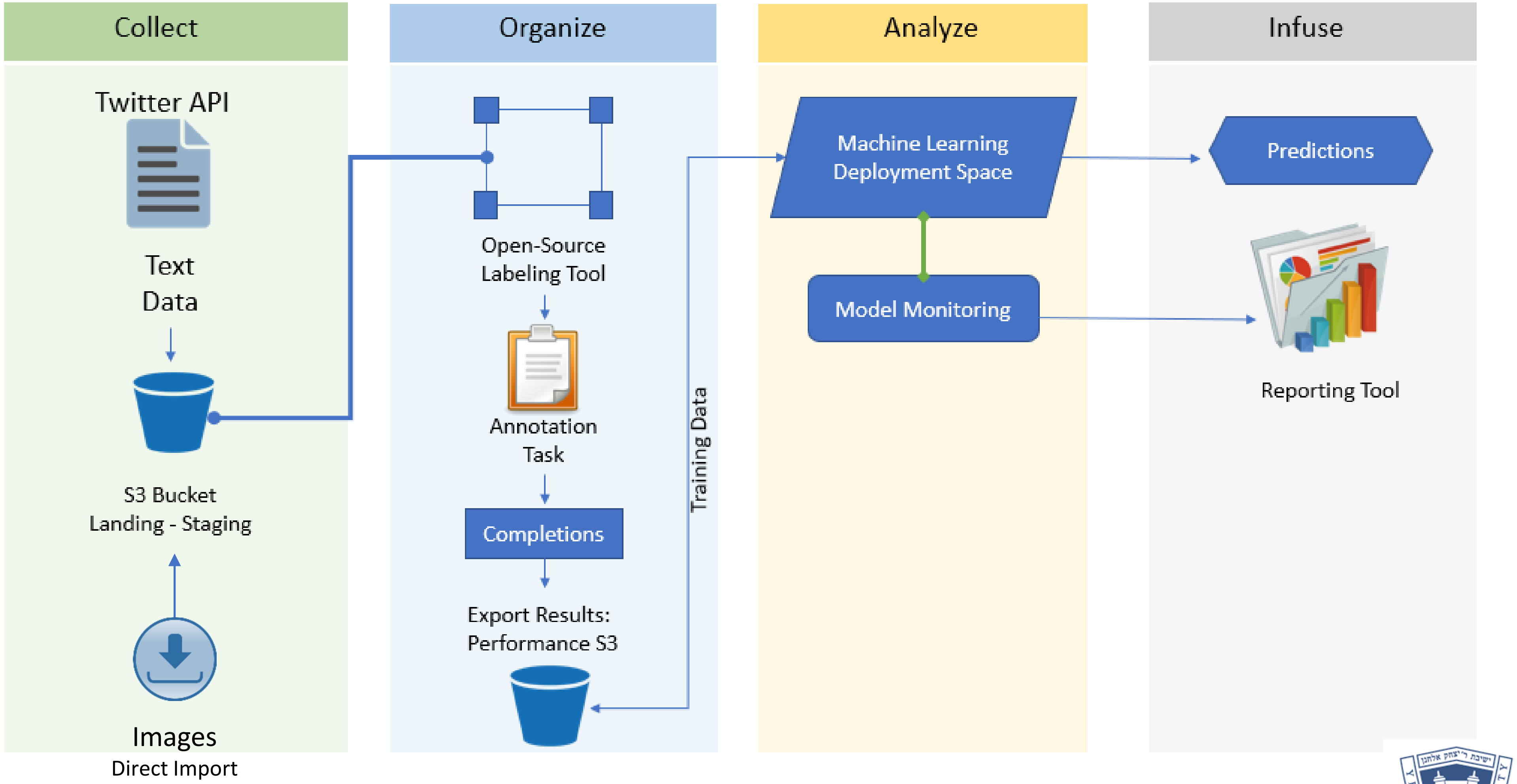
Katz
Katz School of Science and Health

Proposed MVP Application Architecture



Katz
Katz School
of Science and Health

Application Process MVP Development Ladder



Katz
Katz School
of Science and Health

GitHub Project Content

This section will be providing an overview of each file provided in this GitHub Repo.

1. **Annotations**: Contains CSV with sample annotation glossary.

- image_annotations_glossary.csv – file containing the generic sample for an image annotations glossary
- text_annotations_glossary.csv – file containing the generic sample for an image annotations glossary

2. **Documentation**: Contains PDF documentation, including presentations, architecture diagram and logical data flow

- Architecture Diagram Folder: YU Labeling MVP Proposed Architecture.png – file containing proposed architecture of the solution
- Labeling Tool Logic V1.pdf – file containing the V1 logical data flows
- YU Hate Speech Backend Engineering Focused Project Design Proposal V1 2.5.22.pdf – file containing MVP proposal
- YU Labeling Tool Mid-Term Presentation.pdf – file containing the mid-term presentation
- YU Labeling Tool Final Presentation.pdf – file containing the final presentation

3. **Installation**: Contains TXT file with Label Studio installation instructions for Windows OS

- Installing Label Studio on Windows Instructions.txt – file containing Label Studio installation on Windows OS instructions

4. **Scripts – Pipelines** : Contains TXT files with data pipeline ensemble instructions and IPYNB files with Twitter text data mining sample pipelines

- Twitter_Data_Pipeline_V1.ipynb – file containing V1 of Twitter text data mining python script
- Twitter_Data_Pipeline_V2.ipnb – file containing V2 of Twitter text data mining python script
- Pipeline 2 S3 to Label Studio.txt – file containing S3 to Label Studio Pipeline setup instructions
- Pipeline 3 Label Studio Annotations to S3.txt – file containing Label Studio annotation outputs to S3

5. **Scripts – Outputs**: Contains IPYNB outputs parsing sample python script

- Parsing_Output_Data_V1.ipynb – file containing V1 of output parsing sample python scripts – raw data to dataframe

Katz
Katz School
of Science and Health

GitHub Project Content Continuation

This section will be providing an overview of each file provided in this GitHub Repo.

6. **Setup**: Contains TXT files with step-by-step setup and configurations instructions

******The files below show the setup and configuration instructions in a logical sequence.*

- Label Studio on AWS.txt – file contains the Label Studio deployment instructions in AWS leveraging Elastic Beanstalk
- Getting Started with Label Studio.txt – file contains Label Studio “getting started” instructions
- Twitter Developer Acct Setup.TXT – file contains the Twitter Developer Account “getting started” instructions
- YU Labeling Tool Setting Up Labeling Tasks and UI.txt – file contains the Label Studio instructions for setting up a labeling task
- Getting Started Labeling Data.txt – file contains the Label Studio instructions for “getting started” with labeling data
- User Management_Adding Collaborators.txt – file contains the Label Studio instructions for adding collaborators

Katz

Katz School
of Science and Health

Next Steps

- Sprint 2 (ML & AI Team)
- Sprint 3 (User Interface Team)