FALL 2020 – CSE3063 Object Software Design Java Project Iteration #3

Group 10

Description

Data labeling is the process of detecting and tagging data samples to be used in machine learning. In this project we analyze, design and implement a data labeling system in an object oriented manner.

In the third iteration of the project, we added a command line interface for human users. This interface asks for user name and password. If the username and password matches, it uses our new labeling mechanism.

Requirements

- Creating an command line interface
- Supporting multiple human users
- Storing human user credentials
- Asking for username and password in CLI
- Adding functionality of terminating software anytime and keeping state
- Labeling bot users with random mechanism if username and password entered blank
- Adding a new rule based labeling mechanism for human users

Glossary

Term	Definition
Dataset	The collection of labels and instances
Label	An informative tag that can be used in assignments
Instance	A data sample to label
User	Someone who uses labeling mechanism
Labeling Mechanism	Process of assigning labels to instances
Assignment	The allocation of a label or multiple labels to a group of instances.
Random Labeling Mechanism	A label mechanism that choses a random label and assigns to an instance.
Performance Metrics	The collections of statistics that are used to measure behavior, activities and performance of a simulation.
Consistency Check Probability	Probability of labeling a labeled instance by a user.
Configuration File(Config File)	The file that used to configure the parameters and initial settings for program.

Domain Diagram:

HumanLabelingMechanism

Log

Controller

Mechanism

NewLabelingMechanism

PLabel

Storage

RandomLabelingMechanism

PerformanceMetrics

Dataset

Label

UserPerformanceMetric

User

Assignment

DatasetPerformanceMetric

Instance

InstancePerformanceMetric

Percentage

User_Instance

Writer