CACI
EVER VIGILANT

CACI

AIML DATASET DISTILLATION

THE TEAM

Peter Gultom
Computer Science - Data Science
Hometown: Aurora, Colorado
Work Experience: Technical assistance for faculty

Denisha Saviela
Major: Computer Science - Data Science
Hometown: Indonesia
Work Experience: Tutoring Instructor

Duan Nguyen
Major: Computer Science
From: My Tho City, Vietnam
Experience: Python: numpy, C++, Java

Cesar Lopez Hernandez
Major: Computer Science - Data Science
Hometown: Aurora, CO
Experience: Mines ITS MSC Lead Consultant

OVERVIEW

The proposed project aims to investigate and implement dataset distillation methods for two specific tasks: Fashion MNIST classification in computer vision and IMDB Movie Reviews binary sentiment analysis in natural language processing. The purpose of the project is to explore the possibility of distilling large datasets into smaller ones while maintaining the performance of machine learning models.

GOALS AND OBJECTIVES

Objective n° 1

A classification system that attains better than 90% top-1 accuracy on Fashion MNIST test data

Objective n° 2

A distillation method that generates Fashion MNIST training data

- training data of 3000 member
- training data of 600 members

Objective n° 3

A classification system that attains better than 90% top-1 accuracy on IMDB Movie Reviews test data

Objective n° 4

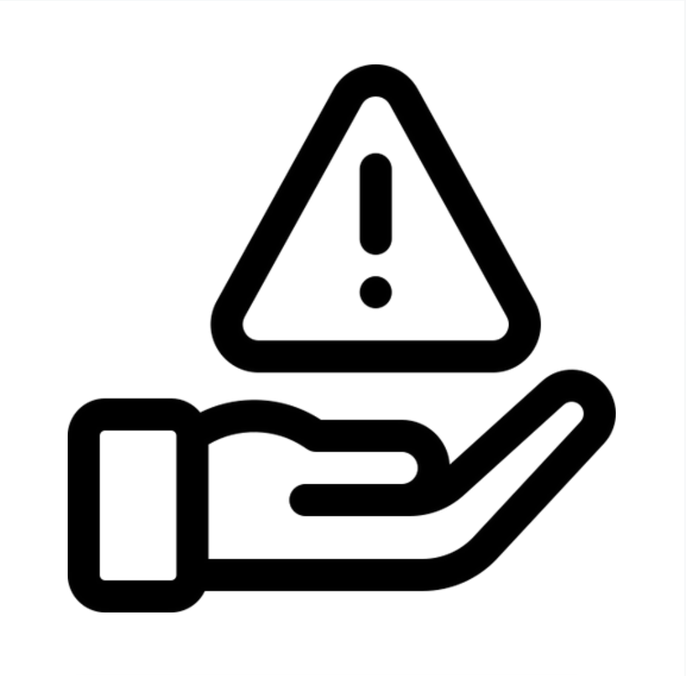
A distillation method that generates IMDB Movie Reviews training data

- training data of 12500 members
- training data of 250 members

RISK

- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.

- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.



DESIGN PROPOSAL

Train full 60000 member FMNIST training set

Attain a better than 90% top-1 accuracy

Generate an FMNIST training set half the size of the original

Generate an FMNIST training set that is one one-hundredth the size of the original

Get at least a 80% accuracy (base steps)