ABSTRACT (DESCRIPTION)

The goal of this project was to use deep learning models for image classification of organic and recyclable images, specifically using neural network libraries including TensorFlow and Keras. Building complex models with the use of specialized architecture was key, to training various neural network models successfully identify and classify whether an image was organic or recyclable. Finally, optimization of model performance and evaluation of metrics (like accuracy) were used to select the best performing model structure and design based on the data analyzed.

DESIGN

Project design consisted of several parts:

1. Preliminary EDA on the image dataset including class distribution, image sizes, shapes, and dimensions
2. Additional image preprocessing including rescaling, resizing, and standardization prior to baseline modeling
3. Performed non-deep learning modeling and a simple baseline neural network before moving onto more complex convolutional neural network modeling
4. Optimization of models and evaluation of model performance with various metrics to identify which model performed the best

DATA

This dataset can be found on Kaggle, here, and contains over 25,000 images. The images were conveniently divided into train and test folders, containing images of two classes: organics and recyclables. Image preprocessing and standardization was required, prior to data transformation for neural network modeling. Roughly 11,000 images were recyclables and 14,000 images were organic images, with approximately 15% of the dataset held out for testing and the remainder for training and validation. Leveraging TensorFlow and Keras, I built a number deep learning models for image classification. The dataset is available upon request, as the files were too large to commit and push onto my Github repo.

ALGORITHMS

Neural Network Models
- Sequential modeling with the use of basic and specialized architecture components including convolutions, max pooling, batch normalization, dense layers, etc

TOOLS

Tensorflow and Keras
- Deep learning machine learning models including a feed forward neural network and more robust convolutional neural network modeling
- Image preprocessing and model optimization libraries

OpenCV
- Computer vision image handling and augmentation

Pandas and NumPy
- EDA, data cleaning and manipulation

Scikit-Learn
- Non-deep learning models (Random Forest Classifier & Logistic Regression) to provide a comparison to deep learning models
- Preprocessing and dimension reduction libraries

Matplotlib, Seaborn
- Image EDA and data visualization, including line and scatter plots

OS, Glob, TQDM
- Misc libraries for file and directory navigation

COMMUNICATION

In addition to the slides and visuals presented, all work is available on my [Github found here](#).