Capstone Project Proposal
By
Ikwuonwu Jonathan Ukaegbu

Domain Background

The capstone project, which is obtained from kaggle - Waste Classification data, uses computer vision to solve a problem in the area of waste management. As a data annotator and mechanical engineering graduate , I am very much interested in completing this project with the use of amazon sagemaker. I grew up in Lagos Mainland and one of the issues we faced was waste disposal and drainage issues, especially during rainy seasons. Waste collection and rubbish disposal plays an important role in the global cleanliness and sustainability drive.Thousands of years ago humans simply dug a hole and buried their refuse but in recent times this won't work because there is a significant increase in the scale of waste being produced, therefore developing a creative solution to solve waste management issues is of great importance.

Problem statement

One of the major ways to solve the problem of waste management is through recycling. Recycling is the process of converting waste materials into new materials and objects. Recycling can prevent the waste of potentially useful materials and reduce the consumption of fresh raw materials, thereby reducing: energy usage, air pollution (from incineration), and water pollution. On kaggle, the project is organized by Sashaank Sekar with the aim of solving recycling issues by reducing the toxic waste ending up in landfills. The waste is segregated into organic and recyclable. I am therefore trying to create a solution that can detect inorganic waste by producing a model to classify waste based on two categories; recyclable and organic waste.

The image is gotten from google

Datasets and inputs

The dataset is made available by Kaggle, assessed by the link https://www.kaggle.com/techsash/waste-classification-data. The data set is divided into train data(85%) and test data(15%), the test and train data set contains two folders named O(orgainic) and R(recyclable). It consists of

train images(22564 images) which are colored images with inconsistent dimensions. The O folders contain images of organic materials which are not recyclable and therefore should be disposed of while the R folders contain images of recyclable materials. To start up the project, I would unzip and upload these files to my s3 bucket to make them available for training with sagemaker. I will conduct some feature engineering where necessary to get better results. The train images would be used for the training job, after which testing is done twice for the 2 test datasets

Solution statement

The proposed solution is to create a deep learning model that is able to accurately identify recyclable materials with a high percentage accuracy by using amazon sagemaker. This solution can enable countries to better solves waste management issues

Benchmark model

As I work on solving this problem, I hope to beat the model accuracy of 90 percent. This target should be achievable with pre-processing of train dataset, hyperparameter tuning and other feature engineering that would be discovered while working on the project.

Evaluation metrics

The approach to evaluate the performance of this project is by getting its accuracy with the confusion matrix. The confusion matrix has 4 categories: true positives, true negatives, false positives, and false negatives which is used to determine the efficiency of the model. My aim is to achieve a high accuracy after completing this project.

Project design

The workflow to approach the solution is as follows: With the right exploratory data analysis, I will explore the data to understand, prepare and clean up the data (if necessary), making it easy to train the model. Since it is an unbalanced dataset, pre-processing techniques such as data augmentation would be done thereby enhancing the model's performance to obtain the desired solution. After specifying the sagemaker environment, I will commence training the model with a RESNet18/32 and a python script as entry point, creating a pytorch estimator, setting hyperparameters(the learning rate and batch size) before calling model.fit(). Sagemaker debugging and profiling will also be done on the model. The model will be finally deployed to an endpoint.

Reference

- Vhandana Bharti, Jaspal Singh. The Importance of Waste Management to Environmental Sanitation: A Review(2018) https://www.researchgate.net/publication/329864740_The_Importance_of_Waste_Management_to_Environmental_Sanitation_A_Review
- Urban Development Series – Knowledge Papers (2012). Global Waste Management Practices pgno x-xi.