Chris Phillips 10/25/2023

Project Part 1:

Small Data Problem Analysis Report

Complete this document and submit it with your project.

Match the scenario with the most appropriate solution and explain your choice

Scenario #1: Travel Planner Problem

A travel planning company asks customers to share pictures of past vacations/holidays so their staff can identify what kind of trips they enjoy. The company offers three basic categories of trips:

- Exploring in the Forest
- Adventure in the Desert
- Relaxing on the Beach

As part of a new online trip planning software, the company is creating an AI bot that will automatically figure out from the uploaded photos which category is likely to be most appealing to the customer. The challenge is the company has fewer than 500 photos that are categorized, and they feel it will be difficult to train a model using such little data.

Scenario #1: Travel Planner Problem Should you use transfer learning or a synthetic data approach to solve this problem? Please explain your answer in a short paragraph containing 3-5 sentences.	I would use synthetic data to generate additional photos to train the model. To supplement the dataset, the company can generate new images with variations in lighting, positions, and distance to increase the size and diversity of the training set. This would make access easier and speed up the data science progress instead of having to acquire new photos for the database.

Scenario #2 Loan Funding Prediction Problem

A loan company has a fairly large dataset that they want to use to train a model that predicts whether or not a loan should be funded. The problem they face is the dataset they are using has a large class imbalance... they don't have enough examples of loans that were denied. This is creating a model that doesn't perform well, particularly for loans that probably should be denied.

Scenario #2: Loan Funding Prediction Problem Should you use transfer learning or a synthetic data approach to solve this problem? Please explain your answer in a short paragraph containing 3-5 sentences.	For the loan dataset, I would use transfer learning, which can handle the imbalanced data classification problem in multiple class data. Essentially, data that has minor classes without enough examples. In this case, pre-trained models from other financial decision areas could be used like with credit cards and debt consolidation or some other sort of lending data. This would help to fill in the gaps in information and increase the financial decision making prediction.