

|  |  |
| --- | --- |
| 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. | Select transfer learning or synthetic data, and explain your choice here.  Use transfer learning.  Select an existing deep learning model already training for image classification of similar desired images, such as sceneries, landscapes, backdrops, etc. Use this existing model in transfer learning by replacing the last layer in the model with a new layer for the classification to the 3 outputs – forest, desert and beach, and re-train with the 500 photos. This should overcome the low number of photos by using common features already identified from the pre-existing model and further fine-tuned with the training on the last layer for the 3 desired image classifications. |

### **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. | Select transfer learning or synthetic data, and explain your choice here.  Use synthetic data.  Financial or specifically loan data are typically tabular and contains personal information that must be protected. Synthetic data can be created for the load denial class by using VAE as it can generate realistic data by learning the existing relationships between the features in the data. As such, the statistical fidelity and feature correlations are maintained in the synthetic data, truly representing the loan denial class. The class can be made balanced for the pooled data, consisting of the original and synthetic data, when used for training, enabled better performance of the predictive model. |