

## **Project/Research Proposal**

### **What is the problem that you are attempting to solve?**

With the increase in data accumulation in healthcare, pharmaceuticals, and many other industries, data provides an excellent benefit for companies by helping them in their decision-making process. However, implementing machine learning algorithms to newly gathered data is a challenge. Building a model from bottom to top takes many resources, such as time and money, which companies might need to have.

In this project, a pre-trained model, tf2-preview/mobilenet\_v2/classification, will be implemented to the cifar10 dataset from TensorFlow, and a new model will be created from the ground up and implemented to the same data. The accuracy scores of both models will be calculated and compared to understand if a pre-trained model works better than a newly created model.

The dataset that will be used consists of images of 10 classes, including airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. One application of this is to create an app that can classify everyday objects and what they are for educational purposes, for students or foreigners visiting a new country.

### **How is your solution valuable?**

If the performance of the pre-trained model exceeds the performance of the model built from the ground up, companies can use this type of approach and apply it to their future datasets, especially for smaller datasets. This type of approach is valuable by cutting the costs and time of training machine learning models.

### **What is your data source, and how will you access it?**

The Cifar10 data will be uploaded from the Tensorflow website. Using:

`X,y =tf.keras.datasets.cifar10.load_data()` in python, data can be uploaded.

Data is composed of 10 classes as mentioned above. Each class is composed of 6000 images, 5000 for training, and 1000 for testing, which brings to 60000 images total. The images are 32x32 pixels. The data can also be accessed using the link: <https://www.cs.toronto.edu/~kriz/cifar.html>  
The file size is 163 MG.

### **What techniques from the program do you anticipate using?**

Exploratory data analysis, preparing and normalizing the dataset, creating and applying CNN models, tuning the parameters, calculating model performance, creating visuals, making inferences, and applying transfer learning.

I might apply other supervised learning algorithms for comparison.

### **What is the biggest challenge that you anticipate facing?**

The main challenge is the processing time the model will take to run. Because the data involved is larger than any other data I worked on before.