**Capstone project proposal**

***Background information:***This project consists of a very high-quality dataset, US license plate images from 50 states, District of Columbia and 5 territories, i.e., total 56 classes for classification. Based on the information provided, all images are originals, no augmented images are present in the dataset. All images are size 128 X 224 X 3 in jpg format. All images have been cropped so the license plate occupies at least 90% of the pixels in any image. This ensures that even simple models will achieve high training, validation and test accuracy. Also included is a csv file so that users can use that to create their own train, validation and test sets if desired.

Dataset source:  
[https://www.kaggle.com/datasets/gpiosenka/us-license-plates-image classification?select=new+plates](https://www.kaggle.com/datasets/gpiosenka/us-license-plates-image%20classification?select=new+plates)

***Criteria for success:***Delivering a model with > 90% accuracy to classify a license plate image into one of different available 56 classes.

***Scope of solution space:***

1. Data wrangling and exploratory data analysis to understand the distribution and patterns of the historic (training set) data

2. Train the model with most of the available data using deep learning (neural networks with batch normalization and weight initialization techniques) vs. supervised learning (random forest classification etc.) algorithm

3. Test the accuracy of the model on the remaining portion (~ 20%) of the historic data

***Constraints within solution space:***

Model accuracy (in classifying the license plate) > 90%. A lesser accurate model could potentially result in wrong identification of the vehicle owner.

***Stakeholders to provide key insight:***N/A

***Key data sources:***[https://www.kaggle.com/datasets/gpiosenka/us-license-plates-image classification?select=new+plates](https://www.kaggle.com/datasets/gpiosenka/us-license-plates-image%20classification?select=new+plates)