**Capstone project proposal**

**Title:**

Insurance policy purchase prediction based on customer’s shopping history

**Problem Statement:**

When a customer reaches out to an auto insurance company seeking for a policy, he/she is presented with many coverage options, which confuses the customer and the company may potentially lose the customer.

**Why this is an interesting problem:**

A well modeled and trained machine learning application could predict the customer needs based on the purchase history and suggest an appropriate insurance policy to the customer. This would decrease the gap between customer and the company.

**What Data is being used to solve the problem:**

For this project, we will be using the data sets provided by All State Insurance company on kaggle:

Kaggle link: <https://www.kaggle.com/c/allstate-purchase-prediction-challenge/data>

Data set are provided in 3 csv files of total size 9 MB -

1. Sample submissions
2. Test and
3. Train

Sample submissions file has two fields customer\_ID and plan.

Test and Train files have (25 different fields which define the customer and policy characteristics).

**Approach to solve the problem:**

* Is this a supervised or unsupervised problem?
* If supervised, is it a classification or regression problem?
* What are you trying to predict?
* What will you use as predictors?

**Final deliverable:**

Build a machine learning model that predicts the policy that a customer will potentially buy.

Stretch goal: Application deployed as a web service with an API and a simple web app to interact with API.

**Computational resources:**

Local machine would be used for development, training and testing the model. And will be using AWS cloud resources for production ready application.