**Capstone Project Ideas**

Saarthak Misri

**1.**

**Problem Statement:**

Handwriting recognition (HWR) is a ability of a computer to interpret intelligible handwritten inputs from various sources such as paper documents, images, touch screens devices etc.

HWR entails Optical character recognition (OCR), which in turn, plays a key role in several sectors such as banking (bank statements and cards), retail (invoices and computerized receipts) and travel (passport/ID number).

Handwritten digit recognition is an important sub problem of OCR and is focal to the project. We will use the MNIST data set.

**Data Source:**

<http://yann.lecun.com/exdb/mnist/>

**Application of course material:**

**Data Access:**

- Setting up a data base on SQL server for accessing the data

**Model Selection:**

- KNN

- Logistic and Multinomial Logistic Regression

- SVM (\*based on time constraint)

- Deep Learning with Kensas

**2.**

**Problem Statement:**

Recommender systems have become increasingly popular in recent years. They are utilized in a many areas such as movies, music, news, books, research articles, search queries, social tags etc. These systems have also become popular among financial service and life insurance providers.

This project deals with the application of recommendation systems to MovieLens data from GroupLens research.

**Data Source:**

https://grouplens.org/datasets/movielens/

**Application of course material:**

**Data Access:**

* Setting up a data base on SQL server.

**Model Selection:**

- Simple Popularity

- Collaborative Filtering

- Jaccard similarity

- Cosine similarity

- Market basket analysis

- Pearson Similarity

* Evaluation Through precision and recall.

**3.**

**Problem Statement:**

Design a predictive model for recommending an airlines to customers traveling during the next 2 months (baseline: month when I start the capstone project) based on the fares.

If a customer is looking to book a flight 60 days in advance, we suggest if the price is reasonable. This suggestion is binary in its outcome - “Book” or “Wait”.

The data source for this project would be the [expedia.com](http://expedia.com) for which we design a web scraper.

**Data Source:**

Expedia API

**Variables we are looking to extract:**

-Origin City

-Destination City

-Departure Date

-Departure Time

-Arrival Time

-Total Fare

-Airway Carrier

-Duration

-Class Type - Economy/Business

-Flight Number

-Hopping - Boolean

-Taken Date - date on which this data was collected

**Application of course material:**

**Data Access:**

* Interacting with APIs

**Model Selection:**

* Random Forest
* SVM
* Deep learning (Keras)