**Newspaper articles summary based on of important words in articles titles and content**

Melissa Ngamini

Thinkful Final Capstone Proposal

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

The project is about giving a brief summary of articles content to the user. I would like the program built to be able to scrape the Internet for articles and give the user a summary of those. There is a lot of publications and articles on the net it can be used to narrow down the type of articles an user is interested in.

**How is your solution valuable?**

The solution is valuable because The user after reading the summary of say articles based on the titles and the content can then be able to identify what articles he would like to read in depth or at least be satisfied with the fact the summary is giving a good enough idea of the articles topic.

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

My data is newspaper articles for different publications online. They can be access online through web scraping using a scraping script on particular online publications. It will be important to get articles with varying type of topics so that our model can be able to give good summary for different types of articles with topic varying from sports, politics, religion, fashion and much more.

**What techniques from the course do you anticipate using?**

I plan to be using Tensorflow and Keras on top of NLP and LSA for the articles contents and titles. I plan to use web scraping to import the Data in Jupyter Notebook. Using python tools, like visualization, exploratory analysis, feature engineering and possibly feature selection. We might also need to use some supervised and unsupervised clustering techniques to put articles with similar topics together and then use convolution and tensor flow to give the summary of the articles content.

**What do you anticipate to be the biggest challenge you'll face?**

I think the biggest challenge will be getting scraping the web for articles needed for our project. I have found on Kaggle a dataset that on newspaper articles that I am working. But later, I will need to scrape the Web for various articles there with the information I need.

**Driving Instructions based on Detection of Traffic Road Signs Shape and Color (2nd Proposed topic)**

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

The project will be similar to a Computer Vision problem where I want to be able to correctly identify traffic road signs from the shape and color of say road signs.
I would like the program built to be able to tell us what the driver is to do after correctly identify say road signs from its shape and colors.

**How is your solution valuable?**

The solution is valuable because it can be used in self-driving cars or even with drivers who have poor eyesight. The program will be able to correctly assess what the driver is to do while approaching a particular road sign. Stop at stop signs, slow down in areas where children might be playing, can't make a U-turn if the road sign is No U Turn and more.

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

My data is traffic road signs and they can be access online through web scraping using a scraping script on Google images using specific search tags. It will be important to get traffic signs pictures in different lights and driving conditions so that our model can be able to correctly identify them. I plan to use tags for Stop signs, Yield signs, Children playing signs, Traffic light ahead signs, U-Turn signs etc.…

**What techniques from the course do you anticipate using?**

I plan to be using Tensorflow and Keras because Neural Networks are really good with image processing. I plan to use web scraping to import the Data in Jupyter Notebook. Using python tools, like visualization, exploratory analysis, feature engineering and possibly feature selection. We might also need to use some supervised and unsupervised clustering techniques to put all similar pictures of traffic road signs together.

**What do you anticipate to be the biggest challenge you'll face?**

I think the biggest challenge will be getting the pictures of all the traffic signs we need. Finding a database with all these pictures will be great but not sure if it exist. So we will need to scrape the Web for the pictures there but the tag may have pictures that will be of no help to us. So we will need to find a way to extract the information we need and reject what is unnecessary.

This was my initial idea. But I couldn't figure out a way to actually do it.

## Proposal For Final Capstone (INITIAL)

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

In this project, we would like to be able to tell say clients what platforms between Facebook (Marketplace), Instagram and Pinterest they will have more chance selling their products.

### How is your solution valuable?

Since Nowadays a lot of people like to use their social media to conduct their business, I though it would be a good product to help newcomer or maybe people with not much of a following on say social media platforms what they need to do to make sure that they can sell their product and make a profit. For particular goods and services, certain social media platforms are a better fit. Wanted to be able to use a model that can help with figuring this out.

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

Most of the data is going to be obtained from scraping the social media platforms mentioned earlier meaning Facebook Marketplace, Pinterest and Instagram shops. Trying on this problem I realized that with the social media platforms having restricted their API access, I was unable to get the data that I needed.

If you have any suggestions on how to access data, I would love to have your input.