Capstone Project Proposal

- ONLINE PURCHASE INTENTION -

By: Tanvi Gandhi

November 12, 2019

- ## MOTIVATION

In this project, we will develop a classification model measuring the user's intention to finalize an online purchase transaction. By analyzing the various behavioral attributes of the user while on a online shopping platform can give important insights to target revenue generating customers.

- ## WHO MIGHT CARE?

Due to the boom of the online e-commerce industry in recent years, our problem stands relevant for various segments who want to monitor user behavior. The clientele in question are the various digital shopping websites/vendors/platforms who would like to maximize their revenue by taking into consideration the user intent by the various actions he performs while on the website. Once model is built, it can be applied in real time so as to focus on those customers more who have a higher probability to purchase a product. For those customers, whose probability is less likely, extensive recommendations can be provided to increase their website time and conversion marketing can be applied to turn them into revenue generating customers. It would also be possible to identify the times of the year when sales increase and thus help in providing better response time to customers. The vendor can focus on the season special products and highlight them during the peak season. There can be various application of the model and more input feature vectors would aid in tailoring a better model.

- ## DATA

The data is extracted from the UCI machine Learning Repository (link: *https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset*). It is a recently donated dataset and is used in the research study by Sakar et al (link: *https://doi.org/10.1007/s00521-018-3523-0*).

The data is collected for different users (12,330 sessions specifically) over a period of 1 year so as to avoid any tendency to a specific campaign, special day, user profile, or period. It contains 18 features vectors like "Revenue", "Product Related", "Product Related Duration", "Bounce Rate", "Exit Rate", "Page Value" etc.

File: online_shoppers_intention.csv

This file consists of various informational values related to a customer behavior in an online shopping session.

APPROACH

As of now, we can't clearly decide on the approach to be followed and can keep on changing as I learn new concepts and methods. However, on an initial look, it seems apt to use a classification model. We can try various models to compare the accuracy. The tentative workflow can be laid out as follows:

- Understanding the data – Building an initial hypothesis
- Cleaning the data
- Data wrangling – Getting a dataset to perform analysis and apply models
- Data visualization
- Statistical analysis and Machine Learning
- Storytelling and Results
- Conclusion and future improvements

DELIVERABLES

- **Jupyter Notebook:** A python notebook which will contain all the code involved in the process.
- **Final Report:** A document to highlight the entire process followed and key takeaways.
- **Presentation and/or Paper:** A presentation for the audience and/or a research paper to highlight the significant insights.