Project Title: Improving Online Retail Sales Using Customer Segmentation
Project Category: Finance & Commerce
Team PJ: Purevmaa Damdinsuren & Jonathan Pangelinan

Motivation: What problem are we tackling?

Thousands of businesses around the world utilize the power of machine learning to increase the efficiency of their marketing strategies. One such way that they use machine learning includes the process of customer segmentation. By identifying meaningful similarities in their consumers, businesses can develop effective marketing campaigns to better understand their customers and target groups that will increase their product sales. Our project goal is to recreate this process for one business in the following ways:

1. Our project aims to determine specific target groups to advertise certain products to so that our chosen company/business can increase their sales count and boost the efficiency of advertising/marketing.
2. Our project also aims to predict which customers are going to purchase a product or not, based on their purchase history.

Method: What machine learning techniques are we planning to apply or improve upon?

1. Clustering for finding which groups of people to best advertise the product to.
2. Classification for predicting a customer is going to purchase after certain marketing.
3. Regression for determining which aspects of a customer's purchase history influence a customer's chances of purchasing the advertised product.

Intended Experiments: What experiments are we planning to run?

- We are planning on experimenting with something similar to the idea of Predictive Analytics where we can use customers' purchase history to predict what they will likely to buy later or what will be good marketing for the certain business to increase their number of sales.

How do you plan to evaluate your machine learning language?

The metrics our project will use to evaluate our machine learning language includes:

- Classification Accuracy
- Confusion Matrix metrics (precision, recall, sensitivity, specificity, etc.)
- Area Under Curve
- F1 Measure
- Root Mean Square Error
- Coefficient of Determination

Links to relevant datasets:

1. Mall Customer Information
2. E-Commerce Data

Note: To develop useful and effective models for running our analyses, we are hoping to find a dataset that has data on sales for a single product rather than multiple. The datasets above are alternative datasets we will work with for now in case we are unable to find such datasets later on.

One example of prior research on the topic: How to Use Predictive Purchase Behavior Modeling to Understand Consumer; In this research, customers were ranked/classified by their likelihood to purchase. Based on the model, the company was able to operate surgical promotion that 'drove incremental margin from customers who were already motivated to buy and incremental revenue from customers who previously felt no incentive to buy.'

Project Milestone

1. **DAP 1: Project design and milestone**
    - Choose a project topic of your interest and design a milestone plan for the project.
    - Search for preliminary datasets to run initial exploratory analysis on.
    - Identify problems and experiments to conduct on the chosen dataset(s).
    - **Due:** Friday, April 17th

2. **DAP 2: Data preprocessing and exploratory data analysis**
    - Preprocess the data and conduct exploratory data analysis.
    - Study the dataset to find appropriate algorithms to use for creating a model.
    - **Due:** Friday, April 24th

3. **DAP 3: Classification**
    - Apply commonly used classification algorithms.
    - Evaluate models based on their accuracy using test-data.
    - Determine which customers are more likely to buy a product and which customers are more likely to not.
    - **Due:** Friday, May 1st

4. **DAP 4: Regression**
    - Develop regression models: linear regression and logistic regression.
    - With regression models, predict which aspects of a customer's purchase history influence a customer's chances of purchasing the advertised product.

- **Due:** Monday, May 11th

5. **DAP 5: Cluster analysis**
    - Conduct cluster analysis
    - Create groups of customers based on their attribute similarities to target as part of a new marketing campaign.
    - **Due:** Monday, May 18th

6. **DAP 6: Neural networks**
    - Design and implement a neural network.
    - **Due:** Monday, June 1st

7. **Model refinement and data analysis report**
    - Finalize the effectiveness of the machine learning model.
    - Improve model performance where possible.
    - Finalize data analysis report for submission.
    - **Due: Friday, June 5th**