Amany Mohammed Al Luhaybi

amluhaybi@uqu.edu.sa

Project 1 Proposal

Project Motivation

The main motivations for this project are to predicts visitor's purchasing intention which will be useful for the business owners of e-commerce.

Question/need:

- Are the visitor's purchasing intentions are increasing during the weekends?
- Are the visitor's purchasing intentions are increasing during the special holidays?
- Are the visitors satisfied with the website and the products available?
- What are the holiday visitors purchased for the most?
- This project is beneficial to predict the visitor's purchasing intentions in general and could be used by the business owners to target their returned customers with advertisements, discounts and special products during their peek purchasing time. Furthermore, give them good statics if their customers never return nor purchasing then there is a need to enhance the website/product.

Data Description:

- **Dataset source:**
  https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset
- **Description of the Dataset:**
  (Sakar, C.O., Polat, S.O., Katircioglu, M. et al. Neural Comput & Applic (2018).)

  The dataset consists of 10 numerical and 8 categorical attributes.

Feature name	Feature description
OperatingSystems	Operating system of the visitor
Browser	Browser of the visitor
Region	Geographic region from which the session has been started by the visitor
TrafficType	Traffic source by which the visitor has arrived at the Web site (e.g., banner, SMS, direct)
VisitorType	Visitor type as "New Visitor," "Returning Visitor," and "Other"
Weekend	Boolean value indicating whether the date of the visit is weekend
Month	Month value of the visit date
Revenue	Class label indicating whether the visit has been finalized with a transaction

- I'm expecting to be working on the following features:

  - ✓ "Product Related", "Product Related Duration" ,Weekend, Month, special_day, the "Bounce Rate", "Exit Rate" and "Page Value" features
  - ✓ The 'Revenue' attribute will be used as the class label with false and true value .

Tools:

- Sklearn for classification with neural network, SVM and logistic regression, as will this library will be used for building a confusion matrix to evaluate the models. Seaborn, matplotlib and pandas

MVP Goal:

- Basic Data Exploration such as:

  - ✓ The shape of the dataset
  - ✓ Head of the dataset
  - ✓ Info of the dataset
  - ✓ Summary of the dataset
  - ✓ Convert some data types if needed
  - ✓ Statistical Insight
  - ✓ Handling missing/ Duplicate values

- Visualizations and answers after building the model to:

  - ✓ Are the visitor's purchasing intentions are increasing during the weekends?