Capstone – 1 - Web Browsing Data Analysis & Modelling (Supervised Learning - Classification)

Approach:

EDA :

Merged two data files on common field like UserID

Cleaned the data

Analyzed Data using UserID , where 0 as taken as Non-registered User and 1 -n taken as Registered User ..

Analyzed data for Non-Registered User Vs Registered User

Extracted percentile of Product Purchase for Non-Registered & registered User & found Top Products in Demand.

Main data Extraction & Calculation for Target Data

Complete Analysis is on grouped data – so I have utilized the Hint provided in Problem Statement

1. Aggregated UserID & Product Purchased and taken summation of Cart value for Grouped product.
2. Calculated Number of time website visited
3. Found Probability = count (Number of time website visited based on TimeStamp) * count (Product Purchased)
4. Calculated Target Variable using Probability * Cart Value

Split the data in Test & Train

Applied Classification Algorithms

Analyzed LogicticRegression Algo and found Accuracy is too low .

Applied Decision Tree Algorithm & found 94% accuracy with which I am submitting the file .