COM6101 Marketing Analytics with Machine Learning

2023-2024 Sem 2 Assignment 1 (20%)

The Hang Seng University of Hong Kong

Due: 10 March 2023 (Sun) 23:50 @Moodle

Q1. Conduct a cross-selling analysis for the grocery transaction dataset (Q1.txt). (5%)

- Each line in the text file is one independent transaction record. The purchased items in each transaction record are separated by the delimiter of commas.
- Perform exploratory data analysis, and briefly discuss important and interesting findings.
- Perform data pre-processing.
- Identify and analyze the most insightful results from the association rule mining using your chosen parameters.
- Suggest and briefly discuss appropriate cross-selling strategies according to your analysis.

Q2. Conduct a customer segmentation for the wholesale distributor dataset (Q2.csv). (5%)

- The dataset contains a list of customers with their spendings on fresh food, milk, grocery, frozen food, detergents & papers and delicatessen in different regions (Beijing, HK or Shenzhen) via offline or online shopping channel.
- Perform exploratory data analysis, and briefly discuss important and interesting findings.
- Perform data pre-processing.
- Identify and analyze the most meaningful and explainable results from the clustering analysis using your chosen parameters.
- Suggest and briefly discuss appropriate sales or marketing strategies for all the customer segments according to your analysis.

Q3. Conduct a customer lifetime value prediction for an insurance company (Q3.csv). (5%)

- The dataset consists of the following attributes:

Attribute	Meaning	Value
id	ID of a customer	Numerical
gender	Gender of the customer	Male or Female
area	Area of the customer	Urban or Rural
qualification	Highest qualification of the customer	Bachelor, High School or Others
income	Annual income of the customer	<=2L, 2L – 5L, 5L – 10L, More than 10L
martial_status	Marital Status of the customer	0: Single, 1: Married

vintage	No. of years since the first policy date	Numerical
claim_amount	Total amount claimed by the customer	Numerical
num_policies	Total no. of policies issued by the customer	1 or More than 1
policy	Active policy of the customer	A, B or C
type_of_policy	Type of active policy of the customer	Silver, Gold or Platinum
cltv	Customer lifetime value	Numerical

- Perform exploratory data analysis, and briefly discuss important and interesting findings.
- Perform data pre-processing.
- Analyze the results from linear regression using your chosen parameters.
- Suggest and briefly discuss appropriate sales or marketing strategies according to your analysis.

Q4. Conduct a customer churn prediction for an e-commerce company (Q4.xlsx). (5%)

- The dataset consists of the following attributes:

Attribute	**Meaning**	**Value**
CustomerID	Unique customer ID	Numerical
Churn	Churn Flag	0: Not churn, 1: Churn
Tenure	Tenure of customer	Numerical
PreferredLoginDevice	Preferred login device of customer	Computer, Mobile Phone, Phone
CityTier	City tier	1, 2 or 3
WarehouseToHome	Distance in between warehouse to home of customer	Numerical
PreferredPaymentMode	Preferred payment method of customer	Cash on Deliery, CC, COD, Credit Card, Debit Card, E wallet, UPI
Gender	Gender of customer	Male or Female
HourSpendOnApp	Number of hours spend on mobile application or website	Numerical
NumberOfDeviceRegistered	Total number of devices registered on particular customer	Numerical
PreferedOrderCat	Preferred order category of customer in last month	Fashion, Grocery, Laptop & Accessory, Mobile, Mobile Phone, Others
SatisfactionScore	Satisfactory score of customer on service	Numerical
MartitalStatus	Marital status of customer	Divorced, Married, Single

NumberOfAddress	Total number of added address on particular customer	Numerical
Complain	Any complaint has been raised in last month	0: No, 1: Yes
OrderAmountHikeFromlastYear	Percentage increases in order from last year	Numerical
CouponUsed	Total number of coupons has been used in last month	Numerical
OrderCount	Total number of orders has been placed in last month	Numerical
DaySinceLastOrder	Day since last order by customer	Numerical
CashbackAmount	Average cashback in last month	Numerical

- Perform exploratory data analysis, and briefly discuss important and interesting findings.
- Perform data pre-processing.
- Analyze the results from decision tree using your chosen parameters.
- Suggest and briefly discuss appropriate customer retention strategies according to your analysis.

Guidelines:

- You are required to write a Python program for each question. The program code must be commented and executable. Your analysis in text must be written in the program (just like how you answered the questions in our lab exercises).
- The program files must be named as:
    - <your student ID>_Q1.ipynb
    - <your student ID>_Q2.ipynb
    - <your student ID>_Q3.ipynb
    - <your student ID>_Q4.ipynb
    - E.g., p123456_Q1.ipynb
- Submit your 4 program files to Moodle. You DO NOT need to submit data files.
- Some grading criteria will be considered in the evaluation (but not limited to): correctness, non-redundant code, appropriate preprocessing carried out, completeness of the program implementation, quality of your analysis and proposed strategies.
- **Any form of plagiarism is strictly prohibited and may result in disciplinary actions.**

- End -