

المادة	عملي هندسة نظم المعلومات
اعداد	م. حلا النعمة
المجموعات	3 أشخاص
تاريخ التسليم	2022/6/11

Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests and spending habits. Companies employing customer segmentation operate under the fact that every customer is different and that their marketing efforts would be better served if they target specific, smaller groups with messages that those consumers would find relevant and lead them to buy something. Companies also hope to gain a deeper understanding of their customers' preferences and needs with the idea of discovering what each segment finds most valuable to more accurately tailor marketing materials toward that segment.

a sales team in a company has classified all customers (customers.csv) into 4 segments (A, B, C, D). Then, they performed segmented outreach and communication for different segment of customers. This strategy has work exceptionally well for them. They plan to use the same strategy on new markets and have identified 282 new potential customers.

You are required to help the manager in both understanding the current segments preferences, and predicting the right group of the new 282 customers by applying the following:

#### **Exploratory Data Analysis (5 marks)**

- Which segment contains more men? plot a chart that visualize gender distribution over segments.
- Which segment contains higher Married customers? And which contains higher single ones? plot a chart that visualize marital status distribution over segments.
- plot a chart that visualize the age histogram, and one that visualize Work Experience
- Which segment contains younger customers? plot a chart that visualize age distribution over segments.

#### **Modeling (hint: use sklearn.ensemble.RandomForestClassifier) (10 marks)**

- Apply needed data preprocessing technics.
- Apply both decision tree and random forest classifiers
- Test and evaluate the two models, which one is better?
- Predict the segments of the 282 new customers in the csv file (new\_customers.csv)
- Which feature is the most feature affecting the segmentation?

#### **Paper: Write a paper that includes: (10 marks)**

- Study objectives
- Data exploration and analysis
- Used methodology
- Experimental Study
- Results comparison and discussion