Week 7 assignment

Team member's details:

Name: Christopher Irvin Ballon Peralta
Email: cballon@uoc.edu
Country: Peru
College/Company: Universitat Oberta de Catalunya
Specialization: Data Science

Name: Neil Erwin Mauricio Villadoma
Email: neil.mauricio@gmail.com
Country: Peru
College/Company: Universidad Internacional de La Rioja
Specialization: Data Science

Problem description

XYZ bank wants to roll out Christmas with personalized offers to their customers. The group up process needs to be automated and can't exceed 5 groups in total as a result.

Business understanding

One important variable to measure the success of the project is assigned category of the client and best product fit for the client list.

Classification, What Christmas offer should we give this customer based on these features: household income, account type, we can give them

Objectives:

- So based on the age, gross income of household, saving account, funds, mortgage etc., tell me which category I should put this client into so that I can offer them either family based, individual based, expensive Christmas offer
- Perform clustering to get the most relevant kind of product for each client, Evaluate N (number of clusters) if it is optimal for the business
- Specify the accuracy of each client prediction

About the data sources, we can utilize holidays dates, Spain population data and the implicit client table given by XYZ bank to do a clustering unsupervised machine learning.

Category(offer) A (top tier clients, low risk, reliable, very high household income)

Category(offer) B (Very mild high household class income

Category(offer) C (mild risk, medium income, reliable)

Category(offer) D (high risk, low income, less reliable)

Category(offer) E (very high risk, low income, highly unreliable

Project lifecycle along with deadline

- Data gathering/frequency (web scrapping,3rd party Appositional Appositional, etc)

- Feature engineering(outlier(noise), missing (NAN) values, imbalance dataset, proper format style, data cleaning(standardized), derived features, statistical analysis, data analysis)

- Feature selection (dimension reduction analysis, Pearson correlation, heat map, extra tree analysis, Pearson features etc

- Model creation (which model? hyper parameter optimization, hyper parameter training)

- Test the model (confusion matrix, Roach score) or back to data gathering or feature engineering

- Production deployment

- Circle CI, Hadoop, aws, time series data, imbalance, outliers, problems face and how I fixed it, 5-6 models (linear, random linear, random,)

Weeks	Responsibilities	
Members	Christopher	Neil
Week7	Problem description Business understanding Project lifecycle along with deadline Data Intake Report Github Repo link	Problem description Business understanding Project lifecycle along with deadline Data Intake Report Github Repo link
Week8	Data understanding	Data understanding
Week9	Data cleansing and transformation done on the data.	Data cleansing and transformation done on the data.
Week10	EDA performed on the data Final Recommendation	EDA performed on the data Final Recommendation
Week 11	EDA presentation for business users	EDA presentation for business users
Week 12	Model Selection Model Building	Model Selection Model Building
Week13	Final Project Report and Code Power point presentation	Final Project Report and Code Power point presentation