Week 7 assignment

Batch code: LISUM01
Submission date: 07/25/2021
Submitted to: Data Glacier
Group Name: Avengers

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: Anye Igwacho
Email: denwarrenigwacho@hotmail.com
Country: Turkey
College: Abdullah Gul University
Specialization: Data Analytics

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 classification.

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,)

19 - 25 July	Problem description, Business understanding, Project lifecycle along with deadline
26 July – 1 Aug	Data understanding, choose what type of data you have got for analysis
2 – 8 Aug	Data cleansing and transformation done on the data
9 – 15 Aug	EDA performed on the data
16 – 22 Aug	EDA Presentation and proposed modeling techniques
23 – 29 Aug	Model Selection and Model Building
30 Aug – 5 Sep	Final Project Report and Code

Weeks	Responsibilities			
Members	Christopher	Vincent	Anye	Neil
Week7	Preprocessing and cleaning, Problem description, Business understanding, Project lifecycle along with deadline	Preprocessing and cleaning, Problem description, Business understanding, Project lifecycle along with deadline	Preprocessing and Project proposal, Problem description, Business understanding, Project lifecycle along with deadline	Preprocessing and Project proposal . Problem description, Business understanding, Project lifecycle along with deadline
Week8 26 July – 1 Aug	Feature selection Data understanding, choose what type of data you have got for analysis	Feature selection EDA performed on the data	Feature selection. EDA performed on the data	Feature selection, EDA performed on the data
Week9	Feature engineering, Data cleansing and transformation done on the data	Feature engineering, Data cleansing and transformation done on the data	Data cleansing and transformation done on the data	. Data cleansing and transformation done on the data
Week10				
Week 11	Model creation and testing, SVM,	Model creation, naïve baiyes,	Result Analysis, KNN	Result Analysis, random forest

Week 12	deployment, Write Report and Harmonize	deployment, Write Report and Harmonize	Write Report and Harmonize	Write Report and Harmonize
Week13	Final Project Report and Code	Final Project Report and Code	Final Project Report and Code	Final Project Report and Code