Content

Tokyo Olympics: Predict the medals tally
"Data Digest June'21 - Bits & Bytes"
Data Science At Work
Flash Card
Course Overview
Reference Materials
Week 1: Linear Regression
Codes and Datasets Week - 1
Ridge and Lasso Regression - Supplementary Material
Mentoring Session Material Download Before Mentoring Session (Week-1)
Week 1 Graded Quiz
Pre-Requisite Week 2
Week 2: Logistic Regression
Codes and Datasets Week - 2
Mentoring Session Material Download Before Mentoring Session (Week-2)
Week 2 Graded Quiz
Week 3: Linear Discriminant Analysis (LDA)
Codes and Datasets Week - 3
Mentoring Session Material Download Before Mentoring Session (Week-3)
Week 3 Graded Quiz
LM2: Practice Assessments
Ongoing activity- Refresher Quiz
Monographs
Solved Example Business Report
PM Project Self Declaration
Predictive modeling Project
- Business Report - Dos & Donts updated new.pdf
- Data to Insights to Reporting.pdf
- Project - Predictive Modeling
- Predictive Modelling Project Problem 1 - FAQs
- Predictive Modelling Project Problem 2 - FAQs

All Notes

Project - Predictive Modeling

Submission type : File Upload **Due Date** : Aug 01, 11:00 PM **Total Score** : 60
Available from : Jul 16, 8:00 AM

Description

Dear Participants,

Please find below the Predictive Modelling project instructions:

You have to submit 2 files :
Answer Report: In this, you need to submit all the answers to all the questions in a sequential manner. Your answer should include detailed explanations & inferences to all the questions. Your report should not be filled with codes. You will be evaluated based on the business report.

Note: In the business report, there should be a proper interpretation of all the tasks performed along with actionable insights. Only the presence of interpretation of the models is not sufficient to be eligible for full marks in each of the criteria mentioned in the rubric. Marks will be deducted wherever inferences are not clearly mentioned. THE REPORT HAS TO BE STRICTLY SUBMITTED IN A PDF/DOC FORMAT. ANY OTHER FORMAT WILL NOT BE CONSIDERED FOR GRADING. 6 Marks are allotted for the "Quality of Business Report".

Jupyter Notebook file: This is a must and will be used for reference while evaluating.
Any assignment found copied/ plagiarized with another person will not be graded and marked as zero.
Please ensure timely submission as a post-deadline assignment will not be accepted.

Problem 1: Linear Regression

You are hired by a company Gem Stones co ltd, which is a cubic zirconia manufacturer. You are provided with the dataset containing the prices and other attributes of almost 27,000 cubic zirconia (which is an inexpensive diamond alternative with many of the same qualities as a diamond). The company is earning different profits on different prize slots. You have to help the company in predicting the price for the stone on the bases of the details given in the dataset so it can distinguish between higher profitable stones and lower profitable stones so as to have better profit share. Also, provide them with the best 5 attributes that are most important.

Data Dictionary:

Variable Name	**Description**
Carat	Carat weight of the cubic zirconia.
Cut	Describe the cut quality of the cubic zirconia. Quality is increasing order Fair, Good, Very Good, Premium, Ideal.
Color	Colour of the cubic zirconia.With D being the best and J the worst.
Clarity	cubic zirconia Clarity refers to the absence of the Inclusions and Blemishes. (In order from Best to Worst, FL = flawless, I1= level 1 inclusion) IF, VVS1, VVS2, VS1, VS2, Sl1, Sl2, I1
Depth	The Height of cubic zirconia, measured from the Culet to the table, divided by its average Girdle Diameter.
Table	The Width of the cubic zirconia's Table expressed as a Percentage of its Average Diameter.
Price	the Price of the cubic zirconia.
X	Length of the cubic zirconia in mm.
Y	Width of the cubic zirconia in mm.
Z	Height of the cubic zirconia in mm.

Dataset for Problem 1: cubic_zirconia.csv

Problem 2: Logistic Regression and LDA

You are hired by a tour and travel agency which deals in selling holiday packages. You are provided details of 872 employees of a company. Among these employees, some opted for the package and some didn't. You have to help the company in predicting whether an employee will opt for the package or not on the basis of the information given in the data set. Also, find out the important factors on the basis of which the company will focus on particular employees to sell their packages.

Dataset for Problem 2: Holiday_Package.csv

Data Dictionary:

Variable Name	**Description**
Holiday_Package	Opted for Holiday Package yes/no?
Salary	Employee salary
age	Age in years
edu	Years of formal education
no_young_children	The number of young children (younger than 7 years)
no_older_children	Number of older children
foreign	foreigner Yes/No

Important Note: Please reflect on all that you have learned while working on this project. This step is critical in cementing all your concepts and closing the loop. Please write down your thoughts here.

All the very best!

Regards,

Program Office

Scoring guide (Rubric) - Predictive Modeling Project

Criteria	Points
1.1. Read the data and do exploratory data analysis. Describe the data briefly. (Check the null values, Data types, shape, EDA). Perform Univariate and Bivariate Analysis.	9
1.2 Impute null values if present, also check for the values which are equal to zero. Do they have any meaning or do we need to change them or drop them? Do you think scaling is necessary in this case?	5
1.3 Encode the data (having string values) for Modelling. Data Split: Split the data into train and test (70:30). Apply Linear regression. Performance Metrics: Check the performance of Predictions on Train and Test sets using Rsquare, RMSE.	7
1.4 Inference: Basis on these predictions, what are the business insights and recommendations.	6
2.1 Data Ingestion: Read the dataset. Do the descriptive statistics and do null value condition check, write an inference on it. Perform Univariate and Bivariate Analysis. Do exploratory data analysis.	7
2.2 Do not scale the data. Encode the data (having string values) for Modelling. Data Split: Split the data into train and test (70:30). Apply Logistic Regression and LDA (linear discriminant analysis).	7
2.3 Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC_AUC score for each model Final Model: Compare Both the models and write inference which model is best/optimized.	7
2.4 Inference: Basis on these predictions, what are the insights and recommendations. Please explain and summarise the various steps performed in this project. There should be proper business interpretation and actionable insights present.	6
https://docs.google.com/forms/d/e/1FAIpQLScKuVyrmTTM7Pboh0IB4YIBUbJp2NrDZcsY4SCRn3ZUkwmLGg/viewform	0
Quality of Business Report	6
Please reflect on all that you learnt and fill this reflection report. You have to copy the link and paste it on the URL bar of your respective browser. https://docs.google.com/forms/d/e/1FAIpQLScKuVyrmTTM7Pboh0IB4YIBUbJp2NrDZcsY4SCRn3ZUkwmLGg/viewform https://docs.google.com/forms/d/e/1FAIpQLScKuVyrmTTM7Pboh0IB4YIBUbJp2NrDZcsY4SCRn3ZUkwmLGg/viewform	0
Points	60

Submit Assignment

Drag your file(s) here or browse for a file to upload

Add comments

Submit Assignment

Previous Next