ASSIGNMENT 4 – MACHINE LEARNING

Name: Course Name: Student ID:

Background and Context

You are a Data Scientist for a telecom company named "O2R2 Mobile". The Product team of the company wants to enable and establish a viable business model to expand the customer base.

A viable business model is a central concept that helps you to understand the existing ways of doing the business and how to change the ways for the benefit of the telecom sector.

One of the ways to expand the customer base is to introduce a new offering of plans.

Currently, there are 5 types of plans the company is offering - Basic, Standard, Deluxe, Super Deluxe, King. Looking at the data of the last year, we observed that 18% of the customers purchased the plans.

The company in the last campaign contacted the customers at random without looking at the available information. However, this time company is now planning to launch a new product i.e. Wellness Plan. Wellness Plan is defined as a Plan that allows the customer to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being by staying connected to their near and dear ones, and wants to harness the available data of existing and potential customers to make the marketing expenditure more efficient.

You as a Data Scientist at "O2R2 Mobile" telecom company have to analyze the customers' data and information to provide recommendations to the Policy and Marketing Team and also build a model to predict the potential customer who is going to purchase the newly introduced telecom plan.

Objective

To predict which customer is more likely to purchase the newly introduced telecom plan.

Data Dictionary

Customer details:

1. CustomerID: Unique customer ID
2. PlanTaken: Whether the customer has purchased the plan or not (0: No, 1: Yes)
3. Age: Age of customer
4. TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
5. CityTier: City tier depends on the development of a city, population, facilities, and living standards. The categories are ordered i.e. Tier 1 > Tier 2 > Tier 3
6. Occupation: Occupation of customer
7. Gender: Gender of customer
8. NumberOfPersons: Total number of persons planning to take the plan with the customer (since these are Friends and Family plans)
9. PreferredServiceStar: Preferred service rating by customer
10. MaritalStatus: Marital status of customer
11. NumberOfUpgrades: Average number of upgrades in a year by customer
12. iPhone: The customer has an iphone or not (0: No, 1: Yes)

ASSIGNMENT 4 – MACHINE LEARNING

Name: Course Name: Student ID:

13. PhoneContract: Whether the customers has a contracted phone or not (0: No, 1: Yes)
14. NumberOfChildren: Total number of children planning to take the plan with the customer
15. Designation: Designation of the customer in the current organization
16. MonthlyIncome: Gross monthly income of the customer

Customer interaction data:

1. PitchSatisfactionScore: Sales pitch satisfaction score
2. PlanPitched: Plan pitched by the salesperson
3. NumberOfFollowups: Total number of follow-ups has been done by the salesperson after the sales pitch
4. DurationOfPitch: Duration of the pitch by a salesperson to the customer

Note:

You should use Logistic Regression w/ Regularization and SVM techniques for this problem and compare the Logistic Regression technique you use for this problem with SVM techniques and interpret/explain the differences. Picking the right modelling technique with an explanation will be marked. Please note SVM Models can take a significantly longer time to run, so if you have time complexity issues then you can avoid gridsearch on these techniques or sample your training data. But your models should be tested on at least 1000 randomly selected data points

Best Practices for Notebook :

- The notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
- The notebook should be run from start to finish in a sequential manner before submission. It is preferable to remove all warnings and errors before submission.
- The notebook should be submitted as an HTML file (.html) and NOT as a .ipynb

Best Practices for Presentation :

Like in real-world projects, the ultimate destination of any project or work is generally an executive or decision-making meeting, where you are supposed to present your solution to the business problem, based on the project/work you have done. The purpose of this presentation is to simulate that kind of experience and to draw the attention of your audience (a business leader like CMO, COO, CFO, or CEO) to the key points of your project, which are

- Business Overview of the problem and solution approach
- Key findings and insights which can drive business decisions
- Model overview and performance summary
- Business recommendations

Please keep the following points in mind while making the presentation:

ASSIGNMENT 4 – MACHINE LEARNING

Name: Course Name: Student ID:

- Focus on explaining the takeaways in an easy-to-understand manner.
- Inclusion of the potential benefits of implementing the solution will give you the edge.
- Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
- Please submit the presentation in PDF format only.

Submission Guidelines :

1. There are two parts to the submission:
    1. A well commented Jupyter notebook [format - .html] run sequentially from start to finish and then converted to .html
    2. A presentation as you would present to the top management/business leaders [format - .pdf] (you have to export/save the .pptx file as .pdf)
2. Any assignment found copied/ plagiarized with other groups will not be graded and awarded zero marks. An ADR will be filed for the academic integrity violation
3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation
4. Kindly refer to the assessment guidelines and **make sure you check the details** of every section to get a better understanding of the expectations in this project.
5. Submission will not be evaluated if,
    1. it is submitted post-deadline, or,
    2. more than 5 files are submitted

Assessment Guidelines :

Please find below the guidelines for assessment

Presentation - 20 pts

Focus on Visualizations, Insights, discussion and conclusion, Spelling, Grammar, Tonality, Brevity etc.

Code Quality - 20 pts

Code runs correctly without errors is commented and readable, is modularized and well structured

Modelling and Results - 30 pts

Data Preprocessing Steps and Explanations
Feature Engineering Steps and Explanation
Logistic Regression w/ Regularization Model Building and Evaluation (Confusion Matrix, AUC) Steps and Explanation
SVM Model Building and Evaluation Steps, Comparison with Logistic Regression and Explanation

Model Tuning - 30 pts

3 rounds of empirical/rationale based Model Tuning for each model, Final Results Achieved, Potential Next Steps