Predictive Analytics Course Project

January 29th, 2025

Context

Allianz is a global financial services and insurance corporation with approximately $1.8 trillion in assets under management. It is renowned for its extensive range of insurance products and investment solutions and is a significant player in the financial services industry worldwide.

You have recently joined the customer engagement team at Allianz. Your role involves collaborating with a select group of advisors dedicated to crafting tailored investment and insurance proposals for a specific segment of Allianz's diverse client base. Your team primarily focuses on retail clients.

As part of your responsibilities, you have access to a historical dataset that includes a representative sample of the type of clients you are expected to engage with. Additionally, your manager will provide a fresh set of client leads each week.

The challenge lies in the volume of leads you receive weekly; the number exceeds your team's capacity. Each lead file contains information on roughly 7,500 clients, but due to resource constraints, you can only reach out to a maximum of 1,500 clients per week.

From your understanding of Allianz's business model, you are aware that the company earns an average profit of 5.0% on the investments made by clients. Historical data suggests that clients typically invest about 22% of their current balances into various financial products. Furthermore, there is a marketing cost associated with each client outreach. The expense of contacting each client is 5 euros.

Your Task

Your ultimate goal is to develop a detailed, step-by-step strategy to assist Allianz in determining which clients should be contacted each week. A key component of your strategy will be the creation of one or more predictive models to aid in this decision-making process.

You are equipped with a historical dataset. This dataset is available in a CSV file named "period_0.csv".

Each week, you will be provided with files containing information on prospective clients. These are the individuals you need to reach out to. The files containing client leads are named "period_x_prediction.csv," where 'x' represents the week number (you will receive five such files over the course of the project).

Following the completion of each week, you will receive the outcomes for the client file you worked on the previous week. These outcomes can be utilized to augment your historical dataset if you choose to do so. The files containing the results for clients are available in the customer service database.

The customer service operations for this project will become operational on February 5th. We will distribute login credentials and submission guidelines as the work groups are formed.

Your first task is due on February 7 at 23:59.

From then on, you will have a weekly deadline each Friday at 23:59.

Submission Plan

Week 1: You need to submit your client file using only descriptive statistics. This will be a business rule submission. You cannot use machine learning.

Week 2: You need to submit your client file using a classification model to predict client engagement. You cannot model the investment amount.

Week 3: You need to submit your client file using a regression model. Your goal this week is to predict the investment amount instead of the investment decision.

Week 4: You need to integrate both classification and regression techniques to optimize client selection, considering both the probability of investment and the expected amount.

Week 5: You should apply all the knowledge gained throughout the course to refine your model and optimize client outreach for maximum expected profit.

Grading

You need to produce a comprehensive report, with a maximum length of 6 A4 pages, using 1.5 line spacing, Times New Roman font size 12pt. The report should detail the model you developed and provide clear recommendations on how to implement it in practice. The 6-page limit includes any cover page.

Along with your final report, you must submit your final campaign file to Moodle. The final campaign file should contain the following fields:

- id, target, probability of purchase, predicted investment, expected profit.

The grading will be based on the quality of your report, the robustness of your model, and the effectiveness of your overall solution. Additionally, 10% of the project's grade will be determined by adherence to the weekly submission deadlines and the relative performance of the final target selection.

```
#01 - id                    - client id
#02 - age                   - client age
#03 - job                   - client job
#04 - marital               - marital status
#05 - preferred_contact     - preferred contact channel
#06 - balance               - current balance in the investment account (in euros)
#07 - loan_house            - the client has a housing loan
#08 - loan_personal         - the client has a personal loan
#09 - n_marketing_contacts - number of contacts before this campaign and for this client
#10 - call_length           - duration of the call to sell the term deposit (in seconds)
#11 - investment            - amount invested in the ETF product
```