Mini project: Leveraging Machine Learning for Revenue Optimization via Strategic Couponing

Introduction

Many customers order only once in an online store. There are a number of reasons why they don't place another order. To counteract this, online retailers use various customer loyalty measures. One measure, for example, is to send coupons some time after an initial order has been placed. The intention is to encourage the customer to make a follow-up purchase. Even though the vouchers don't cost the online retailer anything directly, it is not a good solution to simply send all customers a discount voucher: Some customers would make a follow-up purchase even without a voucher. In this case, a redeemed discount means less revenue for the online retailer. Therefore, it is crucial to devise a more targeted approach for the distribution of these vouchers.

Task

The task at hand involves constructing a predictive model that leverages various features associated with a customer's initial order. The objective is to determine whether a €5.00 voucher should be issued to a specific customer. The model should be designed to predict if a customer will place a follow-up purchase within a 90-day period following their initial purchase. This information is represented by the `target90` variable in the dataset. Each customer who is predicted to not place a subsequent order will be send a voucher.

Empirical analyses by the media retailer have shown that for 25% of the churning customers, the voucher triggers a purchase with an average order value of €10. So if a voucher is sent to a customer who would not actually have bought again, revenue increases by an average of €1.25. Conversely, sending a voucher to a customer who would have placed an order anyway results in a revenue loss equivalent to the voucher value of €5. For customers who don't receive a voucher, there is no impact on revenues.

The model's performance is evaluated based on the expected revenue across all customers in a given dataset. This is computed by considering the model's predictions in conjunction with the associated costs and revenues. It's crucial to note that the model's effectiveness is directly tied to its ability to maximize this expected revenue. Hence, the model should be optimized with this specific goal in mind.

The dataset for training is provided in the `train.csv` file. Detailed descriptions of the features can be found in the `data_dictionary.pdf` file.

Submission and evaluation

Your grade is composed as follows:

- **Jupyter notebook (50%):** Please create a Jupyter notebook that documents all steps you have done. Use markdown cells to guide the reader through your notebook and explain your code. Use comments where needed. Use plots to present your findings.

  Specifically, these are the steps we want to see:

  - You have familiarized yourself with the data, for example, through exploratory data analysis.

  - You have preprocessed the data in a meaningful manner, e.g. by creating new features, dealing with missing values, feature selection and feature encoding.
  - You have trained (at least) one model and tuned its hyperparameters.
  - You have evaluated the performance of your model.
  - You have tried to understand and interpret your model by applying the metrics you learned in the lecture, including global as well as local methods.

- **Model performance (20%):** We compare the expected revenue from your model with the reference revenue that would result if simply all customers were sent a voucher. To do so, we use a holdout test dataset, which is provided in the `test.csv` file.

  Please run your model including all preprocessing steps on the test set. Save the predictions to a new `csv` file named `test_predictions.csv`. Only include the `customernumber` and `prediction` (your model's predictions) columns.

- **Presentation (30%):** Imagine you have to present your model to the marketing executive of the media retailer. The marketing executive doesn't have much time, so you must limit yourself to 10 minutes. In these 10 minutes, you should explain as clearly as possible why (or not) the model should be used by the company. The marketing executive is particularly interested in the rules according to which the model makes decisions and what's the model's benefit in monetary terms. This 10-minutes summary presentation will be followed by 10-minutes Q&A where we ask you questions about your submission.

In case of any questions, please don't hesitate to reach out via the ILIAS forum.

We highly encourage you to work in pairs, however, you can also work alone. Upload your solution as a zipped archive to ILIAS by **January 20, 2025 9:00 a.m.**. This should include the jupyter notebook (run all cells and don't clear the output), the test set predictions (`test_predictions.csv`) as well as your presentation slides as `PDF`. The presentations will take place on **January 30, 2025** during the usual lecture time.

Make sure you comply with the Examination and Assessment Honor Code outlined below.

Examination and Assessment Honor Code

All members of the School of Business and Economics at the University of Tübingen (faculty, students, and alumni) share a commitment to honesty and integrity. In particular, all members follow the standards of scholarship and professionalism in all examinations and assessments.

By submitting this assignment, students agree to comply with this Examination and Assessment Honor Code.

Students who violate this Honor Code are in breach of this agreement and are subject to sanctions imposed by the School of Business and Economics, the University and its responsible bodies (e.g., the board of examiners ("Prüfungsausschuss")).

1. All members of the School of Business and Economics at the University of Tübingen (faculty, students, and alumni) have the obligation to report known violations to the responsible bodies (e.g., the board of examiners ("Prüfungsausschuss") or the Dean of Programs).
2. You must not represent another's work as your own.
3. You must not receive inadmissible assistance of any sort before, during, or after an examination or other forms of course work that is subject to assessment by faculty members.
4. You must not provide inadmissible assistance of any sort before, during, or after an examination or other forms of course work that is subject to assessment by faculty members
5. Violations of this Honor Code will be handled according to the rules and regulations laid out in the rules for this course.