Master's degree in Banking and Financial Regulation

University of Navarra

Universidad de Navarra
School of Economics & Business

Introduction to Modeling evaluation

The concepts related to Data Analytics and Modeling that we have seen during the lectures are going to be evaluated through this exercise. Please, provide a detailed and substantiated answer to all questions. Each one of the main questions contributes in the same proportion to the final grade.

1- You are part of a business team, and your manager is requesting you to analyze the credit cards portfolio. The aim is to understand the portfolio and the bank's current customer mix. Because of security reasons, your business department has no direct access to the bank's databases. Your manager has been able to get a sub-set of data. The data sub-sample has been sent to you as a '.csv' e-mail attachment.

- Argue if this data will be enough to conduct your work?
- What answer would you provide to your manager?

2- As part of the business team, you are overseeing the development of a new autocure model for the collections department. The development team has a first version of the model which, theoretically, displays a remarkably outstanding performance. You all then start a pre-trial process in which you monitor the performance of the model in real time.

During the pre-trial you observe that the model accuracy has significantly dropped, and it is not compatible with the design metrics.

- Would you accept the model?
- What is potentially happening?
- What would you advise to the development team to improve the model?

3- During the lectures we have seen that, generally, we need to numerize features because of technical reasons?

- What are these technical reasons?
- What are the most common numerization techniques in the sector? Could you list some examples of features that usually need numerization?
- Is there some similarity between the sector's recommended numerization techniques and the ones used to fill missing values?

4- You are presented with a set of models from different analytics providers. All models are related to binary classification and are trained with the same data and features. The proposed providers & models are the following ones:

<u>Provider 1</u>

Model 1: Random forest with AUC=0.85.

Model 2: Logistic model with AUC=0.92.

Model 3: Boosted Decision Tress with AUC= 0.78.

<u>Provider 2</u>

Model 1: Random forest with AUC=0.88.

Model 2: Logistic model with AUC=0.88.

Model 3: Boosted Decision Tress with AUC= 0.88.

Provider 3

Model 1: Random forest with AUC=0.87.

Model 2: Logistic model with AUC=0.81.

Model 3: Boosted Decision Tress with AUC= 0.89.

Your manager requests a report of the solutions of each provider. She is, as well, very interested in knowing your opinion on which provider she should endorse when talking to the Chief Risk Officer tomorrow morning.

5- You are in charge of monitoring the risk of the Personal Loans portfolio of the bank. You perform a quarterly review of the models in which you measure the model performance as well as the feature stability (such metrics are typically assessed by the regulator once a year). You observe that the overall AUC is above the regulatory thresholds (AUC > 0.80). You also observe that 6 features out of 32 have a PSI lower than 25%. In one week, you have the next risk committee meeting which is chaired by the Chief Risk Officer (CRO). In the meeting you are expected to present a status report of the model by answering the following questions.

- Is there some issue?
- What is the materiality of the issue? Does it impact our Cost of Risk?
- What could be the origin of the issue?

Furthermore, the CRO will like to know your opinion and reasoning on the following questions:

- Should the model be re-trained? Urgently?
- What advice would you give to the development team?