SE Applied Risk Management - 434616

Session 01: Introduction

Markus G. Schmidt, MSc. FRM

markus.g.schmidt@uibk.ac.at

Winter Term 2023/2024

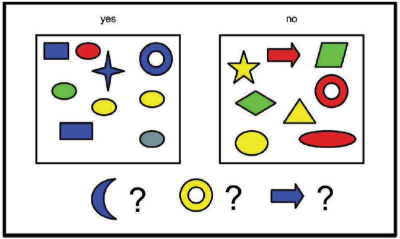
Disclaimer

The views expressed in this slides or communicated throughout the seminar are my own and do not necessarily reflect those of the Bank für Tirol und Vorarlberg AG.

Today

1. The seminar / Syllabus
2. Your case study dataset - A first view
3. Group Formation
4. A small example
5. Group work

Abstraction



(a)

D features (attributes)

N cases

Color	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

(b)

Figure 1.1 Left: Some labeled training examples of colored shapes, along with 3 unlabeled test cases. Right: Representing the training data as an $N \times D$ design matrix. Row i represents the feature vector $\mathbf{x}_i$. The last column is the label, $y_i \in \{0, 1\}$. Based on a figure by Leslie Kaelbling.

Source: Murphy (2012, page 3)

Syllabus — The case study

- You work in a bank's risk management team and are asked by the CRO to develop a new component for the internal rating system. Currently, ratings for corporate clients are based solely on a simple and not empirical based scorecard approach which is biased towards soft-fact information.
- A colleague provides you with a dataset that you can use to develop an empirical model to quantify the default risk of companies. This dataset connects past default events with firm specific balance sheet, P&L and cashflow information. Your direct supervisor expects that the model's output are probabilities of default (PD) estimates which distinguish "good" and "bad" borrowers well.
- The CRO has requested a presentation about your model and your results. Furthermore, your direct supervisor emphasizes that such documents will be relevant for supervisory authorities as well. Hence, the documents should be of high quality and conclusive regarding the expected performance and quality of the new rating system's component.

Excursion: Rating systems

Why are ratings important?

- of course: don't give loans to bad borrowers, but for what else?

Excursion: Rating systems

Why are ratings important?

- of course: don't give loans to bad borrowers, but for what else?

- Risk adjusted pricing
- Monitoring of borrowers
- Management of the loan portfolio
- Calculation of RWAs (Pillar 1) for IRB Banks
- ICAAP (Pillar 2)
- IFRS 9 ECL / Impairment allowances and fair values
- Risk reporting (IFRS 7)
- ...

Objectives and Design

Objectives

- Students get a practical idea about default risk modeling.
- Students are able to work with a dataset and to extract relevant information from the data.
- Students are able to develop a predictive and meaningful scoring model for the assessment of corporations' default risk based on the empirical modeling of corporate default events with a realistic dataset.
- Students are able to communicate their model and results in professionally.

Design

- Lectures provide an overview on how to get started developing a scoring model. Different methods to model corporate defaults empirically will be discussed as well as techniques related to critical aspects of the modeling process (e.g. feature selection, model assessment and selection, missing values, ...).
- Based on the data supplied by Creditreform, students will develop a scoring model and describe their model as well as the results of their analysis in a written report and presentation.

Syllabus — Requirements and Grading

Maximal points achievable by task

- Presentation (30 points)
- Questions after the presentation (25 points)
- Out-of-Sample model performance (25 points)
- Active participation (20 points)
- Extra Points are possible ...

Table: Grading

Total Points achieved	≥ 88	≥ 75	≥ 63	≥ 50	< 50
Grade	1	2	3	4	5

Description of the tasks and requirements - Model performance

- Model performance is evaluated by the out-of-sample discrimination power as measured by the Gini-Coefficient
- Your Gini-Coefficient is compared to the performance of an alternative model developed on the same information
- Tanslated into points P, according to:

$$P = 25 \cdot \begin{cases} \dfrac{Gini}{Gini^* \cdot AF} & \text{, if Gini} > 0 \\ 0 & \text{, if Gini} \leq 0 \end{cases}$$

$Gini^*$ is the alternative model's and $Gini$ is your model's Gini-Coefficient.
AF is an adjustment factor to be defined in the interval $[0.5, 1]$.

- Model sharing across groups *is not* permitted
- Using external data *is* permitted but not expected
- You must provide a valid PD estimate for each firm in the test dataset (i.e. no missing values, interval $[0, 1]$).
- Possibility for point deduction, if your provided estimated PD values are not valid.

Excursion: Gini Coefficient values in practice

Model	Gini Coefficient
Univariate models (individual balance sheet / P&L indicators)	In general, good individual indicators can reach 30-40%, special indicators may reach approx. 55% in selected samples.
Classic rating questionnaire / qualitative systems	Frequently below 50%
Option pricing model	>65% for exchange-listed companies
Multivariate models (discriminant analysis and logit regression)	Practical models with quantitative indicators reach approximately 60-70%.
Multivariate models with quantitative and qualitative factors	Practical models reach approximately 70-80%.
Neural networks	Up to 80% in heavily cleansed samples; however, in practice this value is hardly attanable.

Chart 60: Typical Values Obtained in Practice for the Gini Coefficient as a Measure of Discriminatory Power

Source: OeNB / FMA (2005, page 109)

Description of the tasks and requirements - Presentation and questions

Objective: Convince the CRO with a professional presentation of your approach and model, e.g. by answering the following questions:

- How did you approach the problem?
- What did you do to find a well performing model?
- What are the results of your analysis? How does your selected model work?
- What are the model's variables and how do they influence the model's output?
- What are the limitations of your model?
- Why does you model make economic sense?
- Approximate duration: 25 min

- Question will be directed to the individual group member in most cases, not the whole group
- Each group member has to demonstrate the ability to cope with statistical and economic questions regarding the written report and / or the presentation

Description of the tasks and requirements - Active participation

- I assume that you have an intrinsic motivation for risk management and are not solely interested in the ECTS
- I expect an active participation by all students. Nothing is more boring than silent students in a seminar!

You will have questions - ask!
You will have good ideas - tell them us (and try it out)!
You will want to challenge me - simply do so!
You will want to have a break - let me know!
I will have questions - please answer / comment / say something!

As I see my task

- I want to support you such that you get the most out of the seminar - hence we should work together
- We will cover in the seminar some aspects that are new for you or at least not 100% clear at once. It is important that you inform me about that!
- The final output is important - nevertheless the path to the final output is more important for you. I will like to support you on that path
  - I will try to answer all E-mails in 24h - expect an answer in the night
  - If a meeting outside the sessions is necessary, we will find a possibility to so - mail me.
  - If the question is of general nature or probably of interest to the class in general, please post the question in the discussion board. I will check it regularly (depending how you use the discussion board)

Syllabus — Timeline + Deadlines

Session 01 — Saturday 2023-11-18 — Today

- Group formations, a simple example
- Get familiar with your dataset by getting an overview about the included data and variables. Identify problems within the dataset and think about possible solutions. Think about financial ratios and the design of variables you may want to consider in your model.
- Review the material again and identify aspects you may want to refresh
- Get organized within the groups (Mobile numbers, E-Mails, GitHub, Slack, Dropbox, JIRA, shared calender, ...)
- Getting started with the project

Session 02 — Saturday 2023-11-25 — The next session and project start

- Some additional background
- Q&A with BTV's RMF
- Getting started with the project

Syllabus — Timeline + Deadlines *cont.*

Voluntary submission — Friday 2024-01-06 09:59 — E-Mail

- Send me your intermediate out-of-sample PD-estimates in the correct format. I will inform you about your model's out-of-sample performance as soon as possible.
- Additionally, you can send me any other information about your model etc. to get feedback

Deadline — Saturday 2024-01-20 09:59 — E-Mail

Please send me the following files

- Your final PD-estimates in the correct file-format.
- Your scripts / modified datasets such that i'm able to replicate your results (in a zip-file!)

Session 03 — Saturday 2024-01-27 — Presentations

- It's time for your presentation and and my questions
- Presentation of your out-of-sample model performance

Literature

Statistical (Machine) Learning

Introduction level

- James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). An Introduction to Statistical Learning with Applications in R (7 ed.). Springer. (free pdf: https://www.statlearning.com/)

Intermediate to advanced level

- Hastie, T., R. Tibshirani, and J. Friedman (2009). The elements of statistical learning: Data mining, inference, and prediction (2 ed.). Springer.
- Murphy, K. P. (2012). Machine learning: a probabilistic perspective (1 ed.). MIT Press.

Regulatory (available online)

- EBA (2017): Guideline on PD estimation and LGD estimation and the treatment of defaulted exposures
- ECB (2019): Guide to internal models.
- Federal Reserve (2011): Supervisory Guidance on Model Risk Management.
- OeNB / FMA (2005): Rating Models and Validation.

Your case study dataset - A first view

VariableDescription.xlsx
This excel file contains descriptions of the variables as used in `Training_Dataset.csv` and `Test_Dataset.csv`.

Training_Dataset.csv
You will use this dataset to develop your model.

Test_Dataset.csv
You will use this dataset to calculate the out-of-sample PDs based on your developed model. These PDs are used to evaluate the model's out of sample performance. The file structure is equivalent to the structure of `Training_Dataset.csv` - with the exception that the default indicator variable is not included in the test dataset.

Example_Upload.csv
An example file demonstrating the expected structure of the file (out of sample PDs) you will have to send me. To be more clear, I expect an .csv-file with header, separated by comma (,). The first column must contain the ID of the firm (column name = *id*) and the second column your estimated PDs (column name = *group_x* where *x* is your group number (simple digit)). The name of the file should be *results_group_x.csv* where *x* is again your group number.

Let's form some groups

... and have the first break after that

A small example

Let's have a look at a small example to get started :)

Group work