Data Science & AI in Banking

DI Gernot Griesbacher
Gernot@xGG.at

Machine Learning Graz Meetup
14th January 2020

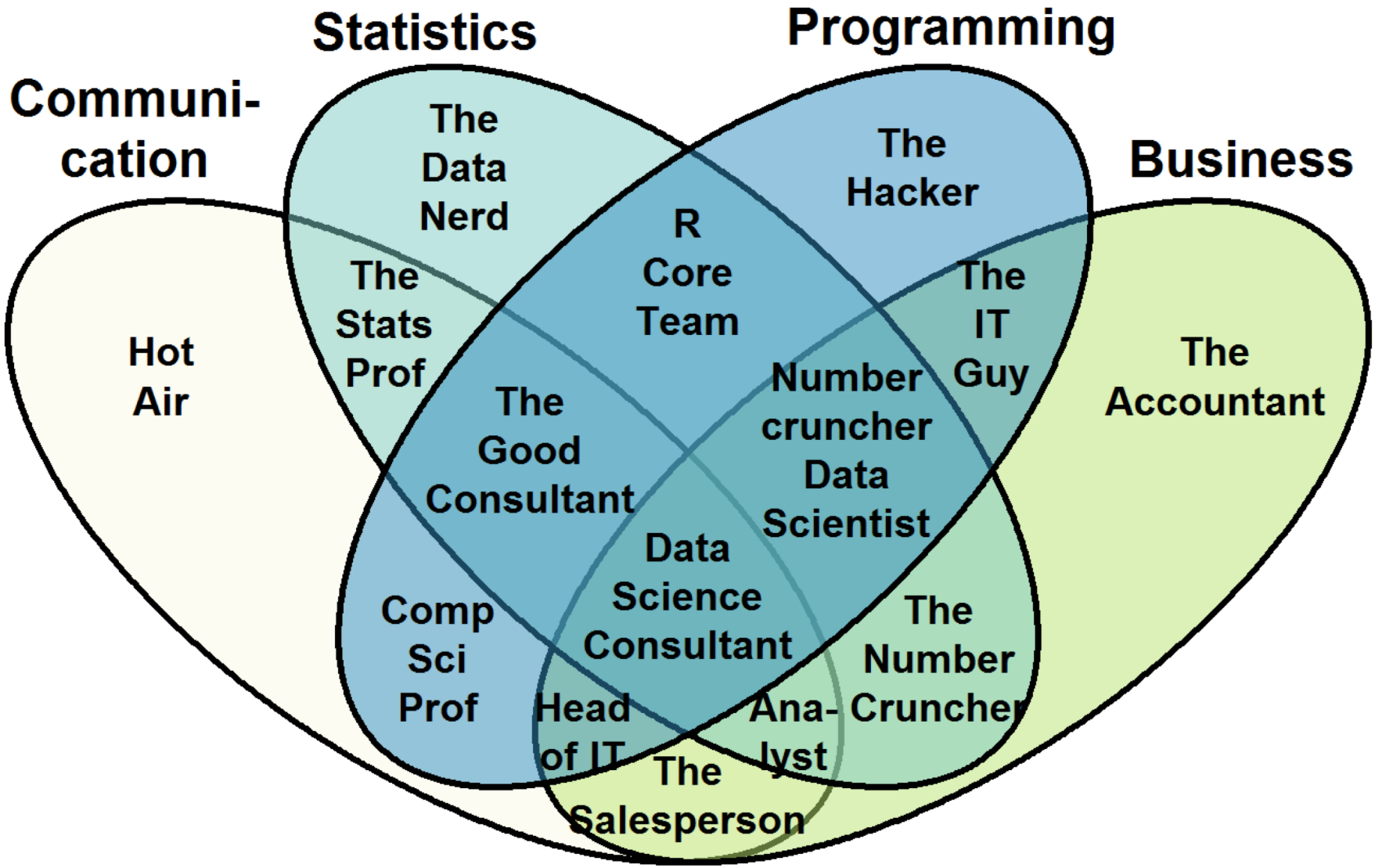
Machine Learning Graz

About Me

- Gernot Griesbacher
- Studied Telematics at TU Graz „Machine Learning“ and „Control Systems“
- Freelancer for software development and consulting since 2008.
- Since 2016 consulting in core-banking:
  - System and data analyst
  - Specialization in risk controlling / rating
  - Projects within central data integration of a bank group

Machine Learning Graz

The Data Scientist Venn Diagram[1]



1 https://datascience.stackexchange.com/questions/2403/data-science-without-knowledge-of-a-specific-topic-is-it-worth-pursuing-as-a-ca

Machine Learning Graz

About my expirience in banking

- "Buzzword"-projects around Machine Learning (ML), artificial intelligence (AI) and "Blockchains"
- In banking there is a high emphasis at the **business domain**, which is organizationally separated from IT domain.
- I would like to explore with you these topics from their "business domain", as programming and statistics is probably your key domain.
- My goal is to give you some ideas, why "modern" IT and Data Driven Design is difficult in Banking & Finance.
    - Disclaimer: "Many statements are my personal opinion based on my experience in Germany and Austria"

Agenda

- Introduction
- IT reality in Banking & Finance
- Risk control as example
- Rating as data user
- Conclusions & Discussion

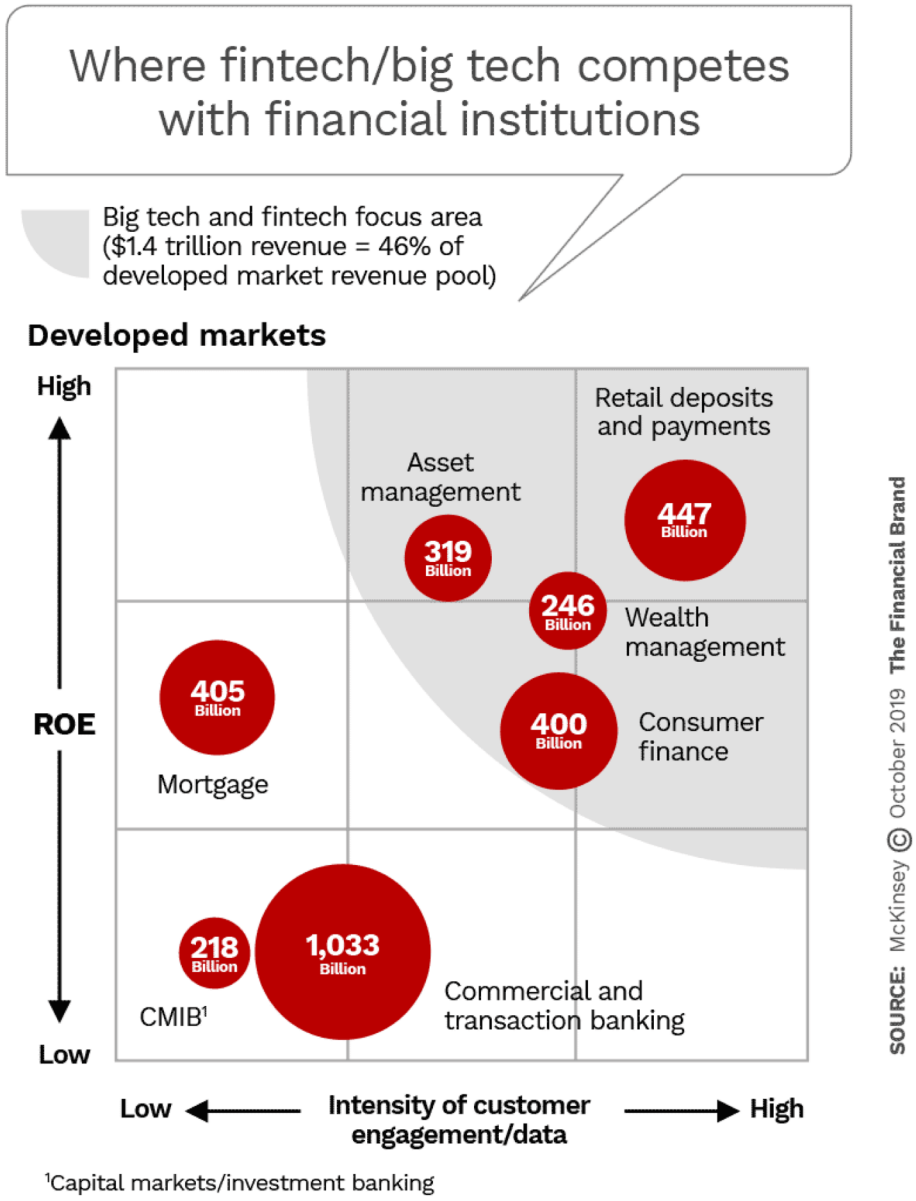
Machine Learning Graz

IT Reality: Banking Markets

There are three main ways banks make money:

1. charging interest for money they lend
2. charging fees for services they provide (e.g. access banc account, bank guarantees)
3. trading financial instruments on the financial markets.

- Currently traditional financial institutions face a lot of competition in the service area



https://thefinancialbrand.com/89741/financial-institution-innovation-fintech-big-tech-cx-ecosystem-merger-partner/where-fintech-big-tech-competes-with-financial-institutions/

Machine Learning Graz

IT Reality: Challenges

- In “traditional” banking the IT has some specific differences to other branches.
- Doing business is important, so banking advisor on the market just care for selling products.
- Financial products are very diverse in nature. There are many completely different ways to model a complex contract.

- New Challenges since 2008 crisis:
  - Risk needs to be measured systematically according to various standards (f.e. BCBS 239, Basel II and III, MaRisk).
  - Laws (CRR, KWG, MiFID, …) for regulatory reporting are complex and are growing more and more in quantity and complexity for reporting obligations.
  - High demands for Compliance, Data Governance and Security. As well as force to open up for new markets (f.e. creation of APIs in online banking with PSD2 requirement)

Machine Learning Graz

IT Reality: Status quo

- IT-systems in banking have some special properties:
  - Central IT infrastructure is mostly pre-PC area (IBM Host-Systems). Organization providing these systems are often “historical grown” with different mindset.
  - There are a multitude of different IT Systems, so that data can’t be easily joined or aggregated. Mostly **non-centralized** data marts provide data sharing within groups.
  - There is currently **no general standard for data modelling** in financial industries.
  - **Data quality problems** because of many manual processes (fax forms) and non-digitized contracts. Many historic data migrations reduces quality further.

- One approach is to collect all data in an **integration platform**, that is independent of specific application formats and serve different data needs.
  - So for each IT system you need to have a loading programs (ETL) which transforms data according to the central data structure in accordance to the business logic.
  - On this platform you can run regulatory processes (e.g. automatic default detection) and serve every department with bank wide consistent data that is approved by business analysts, who take care of many different demands.

Machine Learning Graz

Risk Control as example

- Most of banking business are based on "trust", which can be quantified by the **risk** of a business with a business partner.
- Since 2008 crisis, banking authority enforce regulation, how risk needs to be quantified.

- The banks "trust" in an debtor as much as he is able to pay back the debt. This is measured systematically in the Rating:
  - "A credit rating is an evaluation of the credit risk of a debtor (an individual, a business, company or a government), predicting their ability to pay back the debt, and an implicit forecast of the likelihood of the debtor defaulting"[1]
  - This likelihood is called probability of default (PD)

1. *Kronwald, Christian (2009). Credit Rating and the Impact on Capital Structure. Norderstedt, Germany: Druck und Bingdung. p. 3. ISBN 978-3-640-57549-7.*

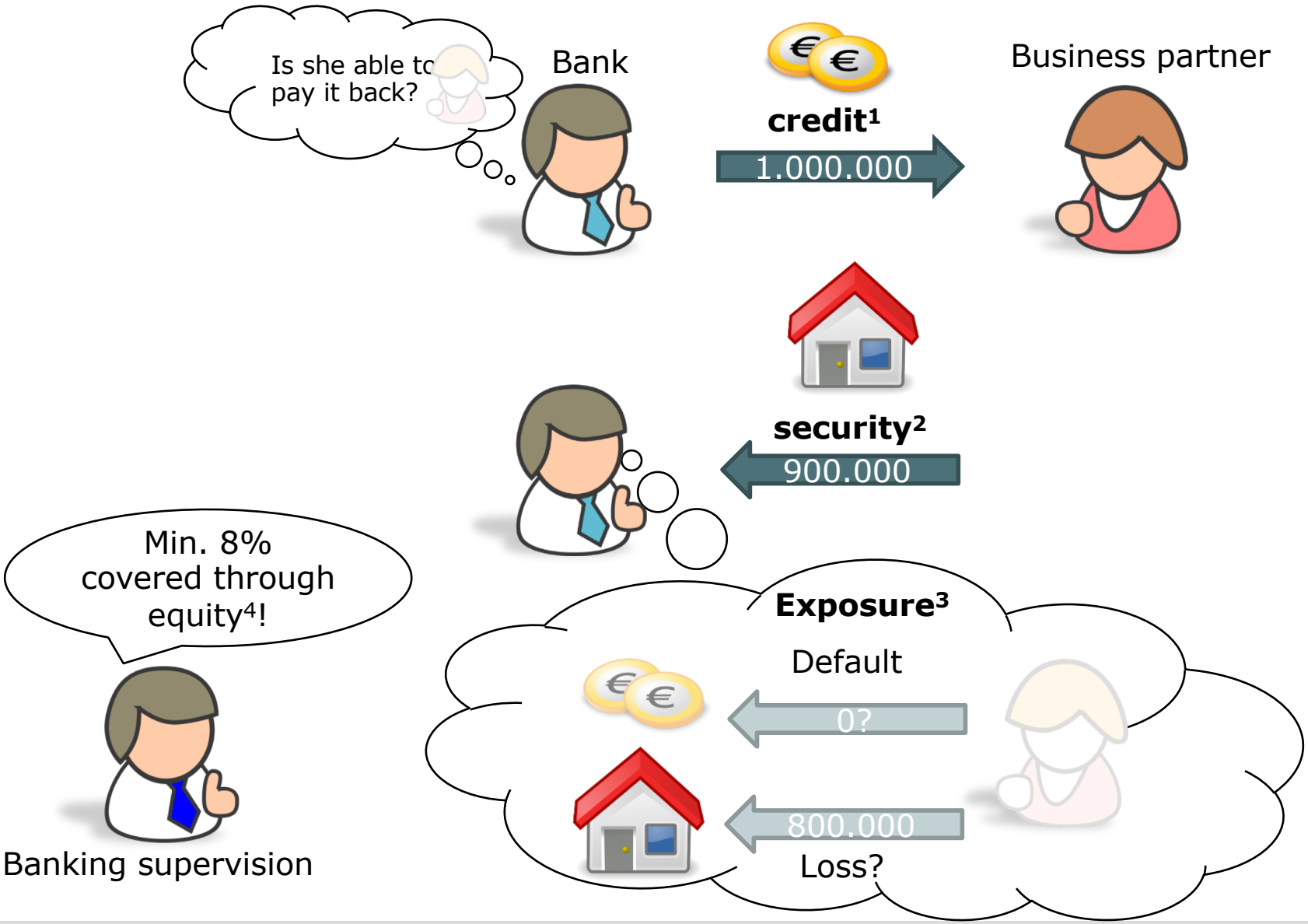
Machine Learning Graz

What is credit risk?

1. Credit: Bank borrows money for a time period to business partners (debtor)
   - But not without further backup – as there is a probability to lose all the money. Creditworthiness of the debtor → this results in a **probability of default (PD)**
2. Security: Deposit of collateral if the business partner cannot repay the money. In picture a house, but could also be other person (bail) or any property of value (money, shares, etc.)
3. Exposure at Default (EaD): What happens if the business partner cannot repay the money?
   - Collateral is used (sold), this should ideally cover the default completely. But this is mostly not the case.
   - In practice: for economic reasons the bank wants to grant more loans than it can back up with collateral.
4. Equity: Bank supervision regulates that the exposure needs to be covered with equity. Not 100% of bank claims covered by bank equity, but only 8% (Basel II guideline) → cost
5. The aim of the bank: to achieve the equity cover as cheaply as possible or to minimize the risk of loss. There are two options:
   - By choosing carefully debtors and valuating them as cheaply as possible
   - By counting as much as possible as security

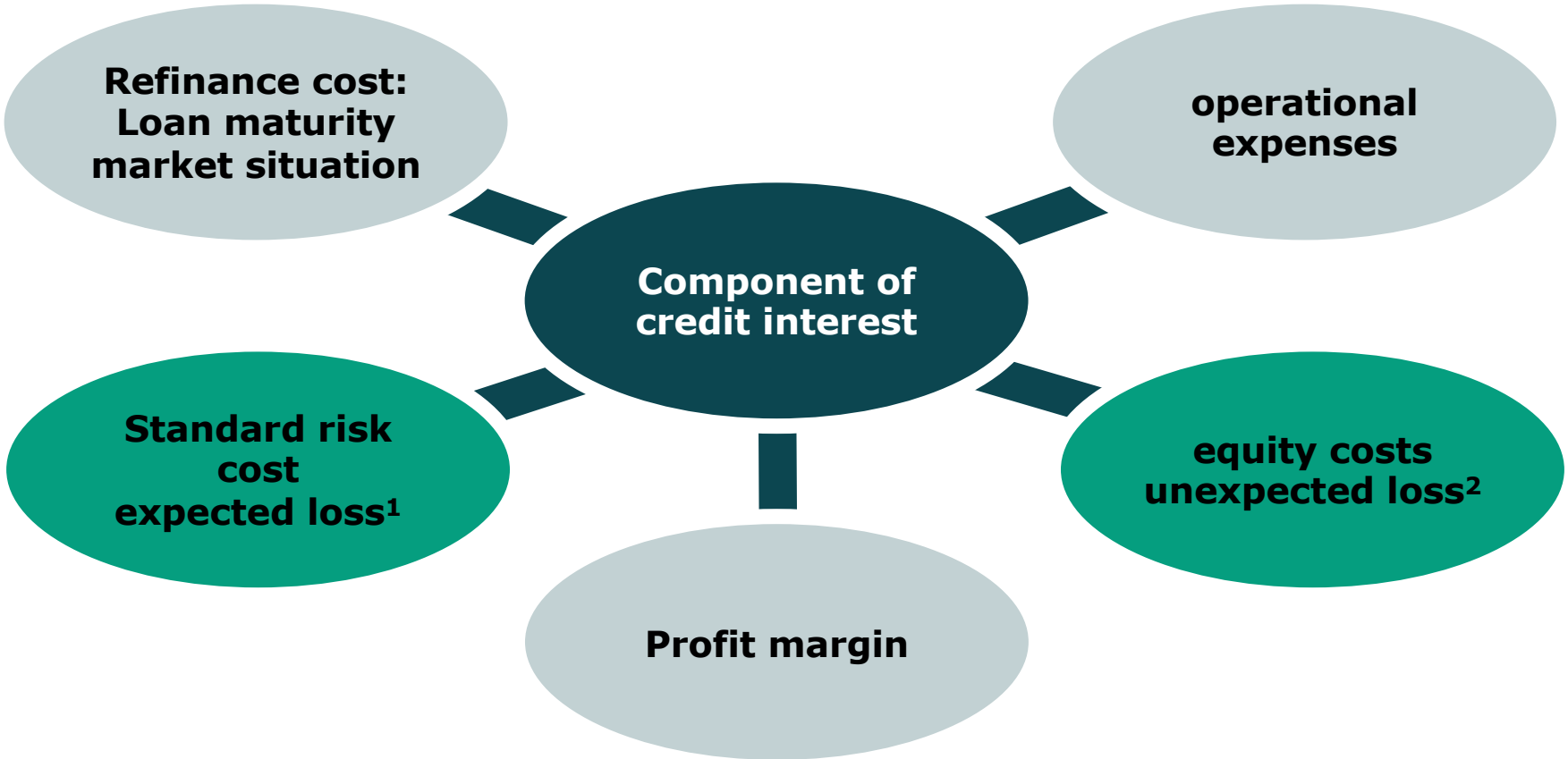
Machine Learning Graz

What is credit risk?



Rating as basis of business model

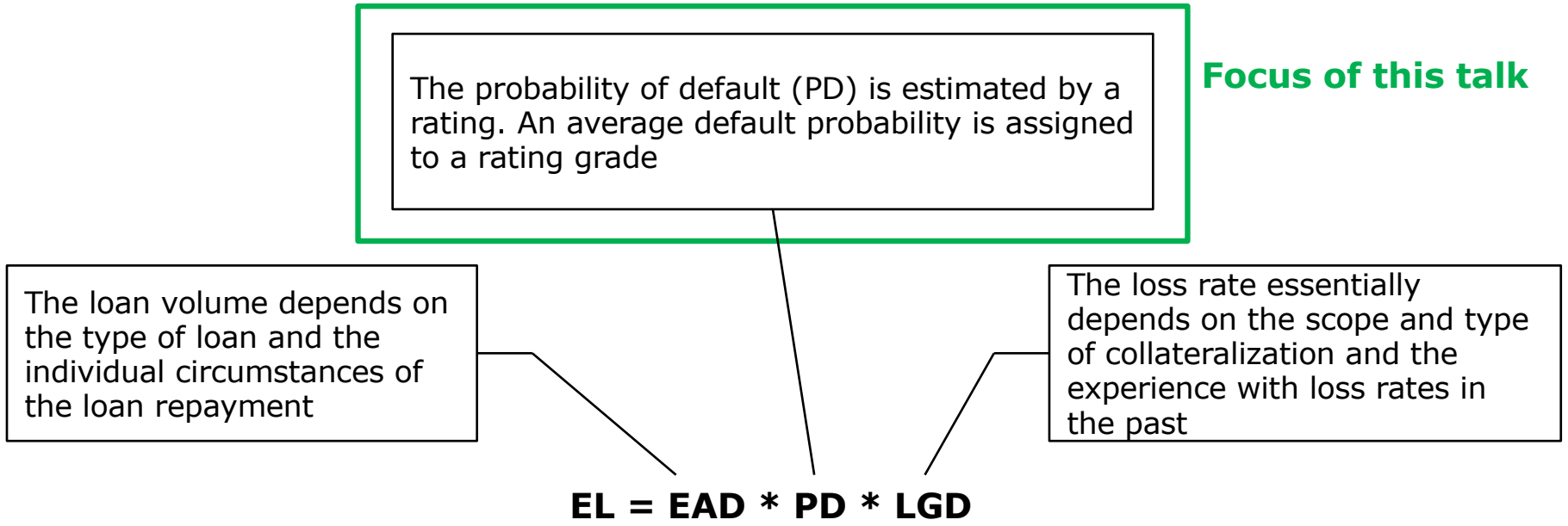
- Credit risk is highly important and essential for the banks business model, as it defines substantially **cost** (1 und 2) and furthermore **business decisions** (are we willing to do this business?).



Machine Learning Graz

The probabilistic model of the expected loss

- The expected loss (EL) is strongly defined by the rating result, which can be "designed" by the bank itself. Other components are fixed:
  - The exposed credit volume at default (EAD) is defined by the credit volume subtracted by the securities.
  - The loss given Default (LGD) is a rate which is calculated by historic events how much money is lost by actually "selling" the securities.



Expected **L**oss = **E**xposure **A**t **D**efault * **P**robability of **D**efault * **L**oss **G**iven **D**efault

Machine Learning Graz

Example for Rating Results

- After a rating result (PD) is calculated from a rating procedure in the range [0 … 1] it is displayed in the form you probably know from the newspapers: 1 A+, 2, etc.
- There are also real time signals (warning signal) or defaults that immediately are processed on top of this probability to react to immediate events.
- For example: client Max M. was rated with a specific scoring procedure to 1 (A-):
  - Very good credit assessment before warning signals
  - There is a warning signal for “account debit return”.
  - There is no default evident

Rating result original	1 (A-)
Rating result after warning signal	14
Rating result combined	14

Machine Learning Graz

Rating types and there implications

Internal Ratings

- Ratings, which are done locally at the bank
- Use in the credit decision, as well as for internal control and reporting
- Each customer with active business is rated

External Ratings

- Ratings conducted by Moody‘s, S&P or Fitch
- Primarily used in financial markets (FM), as well as for internal control and regulatory reporting

IRBA applicable

- Basel II and SolvV defined regulations, procedure is checked by banking authority
- Methodological sovereignty in bank
- Use in reporting possible
- Economic approach brings **realistic equity backing**

Non IRBA applicable

- Not applicable in Reporting
- No regulations from official authorities
- **Higher equity needs** as conservative estimation needed

Emittentenrating

- Counterparty rating

Issue rating

- Rating of securities or shares
- Has priority over issuer rating

Machine Learning Graz

From Data to Rating results

Data

Quantitative characteristics

- Measurable properties, e.g. balance sheet data, monthly income, GDP etc
- A rating system that only requires quantitative characteristics can run automatically and is called Scoring.

Qualitative characteristics

- Immeasurable properties, e.g. management situation, company structure etc.
- Qualitative characteristics require expert estimation in order to be "measurable". Usually score or grade is assigned manually

Solver

- The different characteristics are then weighted and aggregated by a rating system. The result is a rating grade and from this the probability of default (PD) is derived.

Front End

- Output of the rating grade based on a predefined scale and the probability of default. Which is used in risk control and business decisions (f.e. does client get this interest rate).

Machine Learning Graz

Conclusion

1. **Rating algorithm defines what you see as good or bad business**. The data features you use in the rating algorithm are essential.

2. You need to have high **quality** and **quantity of data** about your customer. So the rating system needs to have a lot of data from a lot of different systems.

3. The traditional institutions have decades of data, which they use to analyze which feature are important for which customer segment. These features are weighted in an automatic way in a scoring algorithm or manually evaluated by a rating analyst.

4. If you have a good estimation of the involved risk you can outperform competitions or even use it to automate processes.

Questions?

Thank you for your attention.

Gernot Griesbacher

Gernot@xGG.at