

# Applied Data Science in Fintech

Summerschool Bayonne 2022 | Credit Risk Modeling Workshop



[mario.gellrich@zhaw.ch](mailto:mario.gellrich@zhaw.ch)

Dr. Mario Gellrich

# Program for July 7

Time*	Program: Credit Risk Modeling Workshop
11:00 – 12:00	Introduction, organization of groups, installation of software, brainstorming
12:00 – 13:00	Lunch break
13:00 – 14:00	Working in groups on the tasks; technical sessions
14:00 – 15:00	Working in groups on the tasks; technical sessions
15:00 – 16:30	Presentation of results (ca. 10 min. per group)

\*additional breaks are included

# Workshop procedure and philosophy



Image Credit: <https://revistaempresarial.com>

- No permanent presentations by the lecturer
- Group work (8 groups each with 3 students)
- Groups are mixed (BSc/MSc OR Beginners/Professionals)
- Each «expert» group is responsible for one «topic»
- Methods are explained in detail in «technical sessions»
- Students ask members of «expert groups» for help.
- The more experienced students help the others.
- At the end of the day, groups present their results.
- Presentations are evaluated with pass/fail.

# Prerequisites for July 7

## Software:

- Anaconda (<https://www.anaconda.com>)
  - Python 3
  - Jupyter Notebook, Jupyter Lab, Google Colab etc.

## Account:

- Amazon Web Services (AWS)
- GitHub (please provide your user-name here: [Link](#))

## Material for exercises

- Jupyter Notebook (will be used to guide the exercises and provided by the end of the day)

# Definition

*“Credit risk, counterparty risk or default risk is a term used in the banking industry to refer to the risk that a borrower is unable or unwilling to repay the loans granted to it in full or in accordance with the contract. In general, credit risk is the most significant type of risk for credit institutions.”*

Source: Wikipedia

# Available Data

The original data set contains 1000 persons with their attributes and a credit risk classification and is described under this Link: <https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29>

In the exercises, we will work with a subset containing the following variables:

Risk	(categorical <b>target variable</b> : the credit risk of each person in the data set is classified as low or high)
Age	(numerical, age of borrowers)
Sex	(categorical: male, female)
Job	(numerical: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled)
Housing	(categorical: own, rent, for free)
Saving account	(categorical: little, moderate, quite rich, rich)
Checking account	(categorical: little, moderate, quite rich, rich)
Credit amount	(numerical: national currency)
Duration	(numerical: in month)
Purpose	(categorical: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others)

# Modeling method

	<i>Supervised Learning</i>	<i>Unsupervised Learning</i>
<i>Discrete</i>	classification or categorization	clustering
<i>Continuous</i>	regression	dimensionality reduction

Due to the categorical (discrete) nature of the target variable (credit risk), and because there are labels available (low/high credit risk), classification is a useful modeling approach.

Which methods (classifiers) are appropriate?

# Keyword pinboard (basis to define group tasks)

## (Topic 1) Working with Python and Jupyter Notebook

- 
- 
- 
- 
- 
- 
- 

## (Topic 2) Explaining the data and working with data frames

- 
- 
- 
- 
- 
- 
- 

## (Topic 3) Exploratory Data Analysis

- 
- 
- 
- 
- 
- 
- 

## (Topic 4) Data Engineering, Preprocessing

- 
- 
- 
- 
- 
- 
-



# Keyword pinboard (basis to define group tasks)

## (Topic 5) Modeling (LogReg, KNN, RandomForest)

- 
- 
- 
- 
- 
- 
- 
- 

## (Topic 6) Modeling (CT, XGBoost, MLP)

- 
- 
- 
- 
- 
- 
- 
- 

## (Topic 7) Model predictions and -performance

- 
- 
- 
- 
- 
- 
- 
- 

## (Topic 8) Model deployment with flask and Python

- 
- 
- 
- 
- 
- 
- 
-

# Expert groups and definition of tasks

## Expert Group (X): Working with Python and Jupyter Notebooks

- Installation (Anaconda)
- Creating a Conda environment
- Making Python Code running
- HTML formatting (Markdowns)
- Getting help for the functions
- Installing additional Python libraries from within a Jupyter Notebook
- Working with short cuts
- Saving Results as HTML-File
- Jupyter Notebook, Jupyter Lab, Google Collab
- Running locally developed Jupyter Notebooks with Amazon SageMaker

Example!

# Recommended structure of the «expert» sessions

Questions which help to explain the software, functions, methods, etc. by each expert group

Q1: What is ...?

Q2: What problem is solved with ...?

Q3: How does ... work?

Q4: What are typical applications for ...?

# Minimal structure of the final presentation (Jupyter Notebook / HTML)

## 1. Introduction

1.1 Background

1.2 Problem

1.3 Objectives

1.4 Research Question

## 2. Materials and Methods

2.1 Data sources and data description

2.2 Data preparation

2.2 Exploratory data analysis (EDA)

2.3 Modeling credit risk

## 3. Results & Discussion

## 4. Conclusions