Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

SBD3 | **Find out your future salary** (Group Work 1)

Prof. Dr. Branka Hadji Misheva

Predictive modelling project – Overview

Work on an empirical case study and derive insights from real survey data on what determines people's wages

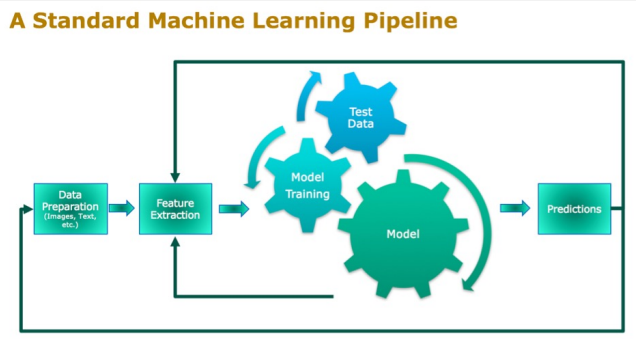
Data Science Survey Challenge

- Industry wide survey of the data science and machine learning community
- An overview of the industry on an aggregate scale
- The data contains information on people's demographics, education, experience and use of programming as well as their current salaries

Predictive modelling project- Briefing

Group Work

- Try to define very clearly what you will do and it what order
- Invest time to really understand the dataset
- Involve all (!) team members in the explorative process
- Consider what are the steps of carrying our a machine learning project



- Apply supervised learning methods that we covered in class

Predictive modelling project– Briefing

- The goal of the work is to use the data "**data_wage.RData**" and build a predictive model of **wages.**

- **Dataset overview:**
  - Subset from a real survey data
  - Dataset dimensions: 10,809 responses across 78 questions
  - Description of the individual variables is included in the file "descriptions_variables.xlsx".

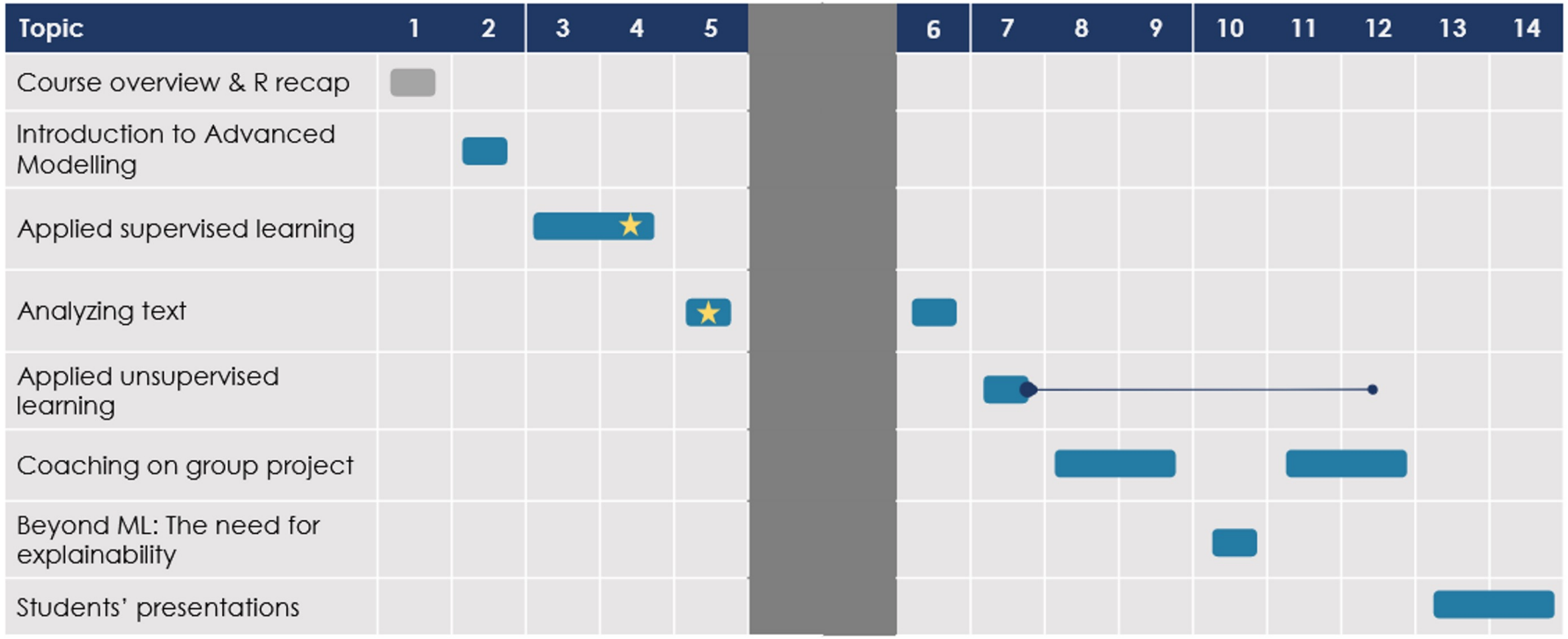
Predictive modelling project– Briefing

- **Objectives of the work:**
  - **Build a model that will predict the variable "wage".** Put differently, the dependent variable of the model should be y = "wage".
  - **You are free to decide** on the features you include as explanatory variables as well as the specific models you use. You need to provide details as to decisions you have taken.
  - Once you have built your predictive model, **you need to explain it (apply explainability models)**! Identify the main features that drive the wage predictions in your model.
  - **How will your model predict for you?** Use your model to predict the wage that each one from the team will earn in the future ☺

  - **Coaching: 23th of April & 14th of May**

Course overview

- Homework handout
- Group project handout



Notes:

- Easter break (no classes): 26th of March and 2nd of April

Evaluation criteria

- **Completeness**
  - You are required to cover all points outlined in the previous slide (under Objectives of the Work), i.e. feature selection, ML training and testing, explainability (including local explainations for your team members).
- **Reproducibility** of results (Script runs without errors, comments help to understand what you have done and why)
- **Appropriateness of methodological approach and quality of result discussion**
  - Throughout the work, you are required to explain your choices and interpret results using arguments from the lecture slides and discussions in class.
- **Design of slides** and presentation style (e.g., argumentation consistency & focus, data visualizations, etc.)
- Ability to **respond to questions**

Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Good luck!