Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

DATA DRIVEN
DECISIONS

Exercise 4

Practical Data Science (PDS)

Agenda

1. General Information of Assignment Solutions
2. Updated Schedule & Group Project Organization
3. Assignment 2: Introduction to Machine Learning

Assignments – General Information

- Solutions are published in GitHub repo: **pds2425/course/assignments**
- Grading Criteria: finish all **mandatory** exercises + notebook is **runnable**
- Assignment 1 overview: most people passed (25/26)
- In case you don’t have the grade, contact me via email: the-viet.nguyen@uni-wuerzburg.de (or discord)
- Important Update: We have 4 main assignments + bonus assignment (participation on discord/wuecampus)
  - You need to complete **3 assignments** to work on the final project
  - You need **4 assignments** to get the bonus
  - Examples:
    - 2 main + participation = eligible for final project
    - 3 main = eligible for final project
    - 3 main + participation = eligible for final project + 0.3 bonus

New Schedule & Group Project Organization

- Group Selection: will be open this week
  - **Group of 3 –** Deadline is 25.11.2024, 11:59 pm

- Project Overview: fine-tuning large language models (LLMs) for question & answering tasks (briefly)
  - Collaboration with snapADDY – more about the topics in 02.12.2024
  - Important lectures:
    - Introduction to Natural Language Processing (NLP) (25.11.2024)
    - NLP Models on Hugging Face (02.12.2024)
    - Prompt “Engineering” (09.12.2024)

Assignment 2 - Overview

- **Predicting Video Game Sales:** Machine Learning + Feature Engineering lectures

1. Pre-processing data
   - Splitting features and target variable
   - Data splitting: train set vs. validation set (no need for a third set for now)
   - Removing missing values (data imputation)
   - Feature engineering: encode categorical variables

2. Training machine learning models
   - Build a Decision Tree & Random Forest
   - Compare in-sample and out-of-sample performance of both models

- **Deadline**: 23.11.2024 at 23:59 pm