

BDA 2021/22

Mid-term feedbacks

General Feedbacks

- Please, from the next presentation you **MUST** stay within the time limit of **20 minutes**.
 - from next mid term on, you will be evaluated also on your compliance with these time limits
- If you have not yet formalized your analytical proposal (or if you are still undecided), write us an email by this week
luca.pappalardo@isti.cnr.it
giuliano.cornacchia@phd.unipi.it

Team Feedbacks

First of all, **great job everyone!**

Overall, all the analysis were performed in an **excellent** way.

Now, some feedbacks to further increase the quality of your projects!

Feedbacks for every team

- Repeat the analysis with/without the new features;
- Use a random seed to guarantee the reproducibility of your analysis;
- update your notebook with the aforementioned suggestions and upload it by next mid term.

WeMine

Dataset: Ethereum Fraud

Feedbacks:

- Inspect the class of outliers;
- Provide a formal definition of your task;
- Comment all the choices in the notebook.

BigData-Islanders

Dataset: Mushroom Classification

Feedbacks:

- Characterize the discarded attributes in terms of class they belong to, to understand if we are removing records of one class predominantly;
- Find stronger association rules (even with lower support), e.g., if the mushroom has some features then it is poisonous;
- Comment all the choices in the notebook.

cpu_in_flames

Dataset: Rental Bike

Feedbacks:

- Use scikit-mobility [1] to partition the city in tiles;
- Correct the 366 days “issue”;
- Select carefully the temporal and spatial granularity for your predictor.

[1] <https://github.com/scikit-mobility/scikit-mobility>

SLM

Dataset: No-Show Appointments

Feedbacks:

- Repeat the analysis with/without the new features;
- Focus properly on the **recall** for your predictor.

How I Met Your Big Data

Dataset: Police Killings

Feedbacks:

- (Optionally) Explore the use of a tool to infer the gender (where missing) from the name (e.g., see [1]);
- Select the final task for your project; actually an interesting option can be to predict the **Dispositions_exclusions** attribute given a police killing (an AI-judge, is it biased?).

[1] <https://pypi.org/project/gender-guesser/>

The Missing Values

Dataset: The Movies Dataset

Feedbacks:

- Make clear how you intend to validate your recommendation system.

Important dates

Mid-term 2: November 17th and November 19th

Papers presentation: December 1st, 3rd, 10th

Mid-term 3: December 15th and December 17th