ABOUT ME

Alberto Danese

Head of Data Science

nexi

- **Professional Experiences** nexi, Cerved, EY, between
- **Education** POLITECNICO MILANO 1863, UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
- **More...**

Competitions **Grandmaster** on kaggle

A²D *allaboutdata.substack.com*

Speaker at AWS Re:Invent, Google, Codemotion, Kaggle and other data & tech events

PROJECT – WHY, HOW AND WHAT

Understanding **Ethics for AI** is key for developing algorithms in a responsible and conscious way, as well as for choosing the right data – the tricky part is translating theoretical notions to a real-world scenario!

We'll deal with **real data** and an actual **machine learning task,** in order to get a practical experience on (some of) the ethical problems that **may arise**

So what's **needed**?

DATA SOURCES	***DATA SCIENCE TOOLS***	***THE HUMAN FACTOR (I.E. YOU)***
kaggle	python, jupyter, SHAP, pandas, dmlc XGBoost, plotly	

THE RULES AND TIMELINE

1. **Follow the outlined analytical path** (in the provided Colab template) but feel free to provide additional analysis and considerations based on your own sensitivity. Focus on the **reasoning**, but use the **code** to analyse and understand the data
2. **Group work:** 5 students (at least one **confident** in Python and computer science)
3. **Evaluation:** focus on the **ethical** part, not on the technical one
4. **Deliverable**: ~10 pages on ppt (in <u>addition</u>, you can <u>optionally</u> provide a Jupyter notebook made on Colab, based on the *provided template*)
5. **Deadline:** 2 november 2022

ABOUT KAGGLE.COM

Leading platform for **machine learning competitions** since 2010

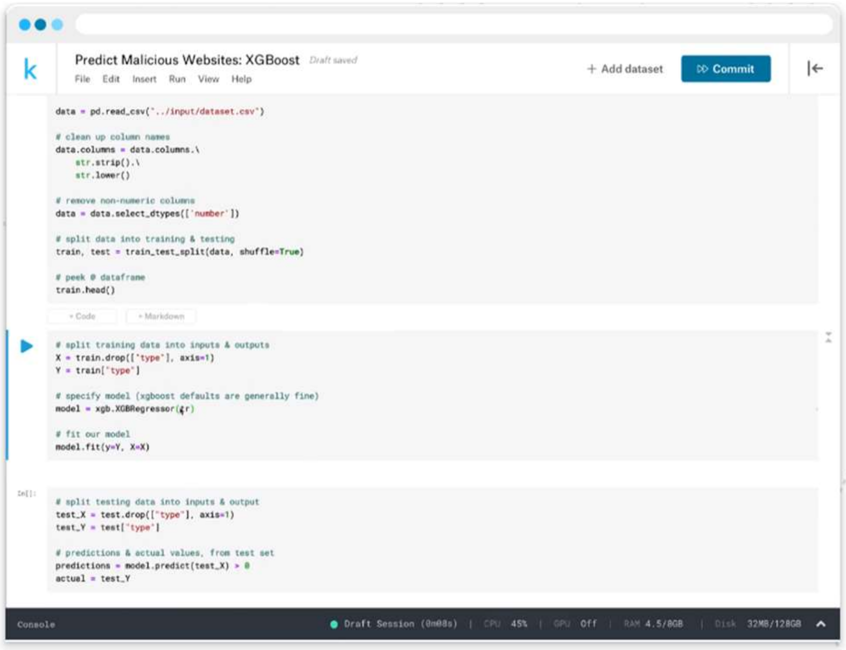
Companies post **real data** and problems that can be solved with predictive modeling / machine learning / AI / some kind of magic!

Data scientists from **all over the world** compete to produce the best algorithms and earn prizes (**15M$ awarded** so far… really!)

Acquired by **Google** in 2017

Grown to a **complete ML platform** with learning modules, code sharing features (kernels), job board, datasets and more

kaggle



OUR PLAYGROUND (1/3)

We’ll use a competition dataset… but not for competing ☺

Task: predicting the risk of not paying a loan (binary classification, metric: AUC)

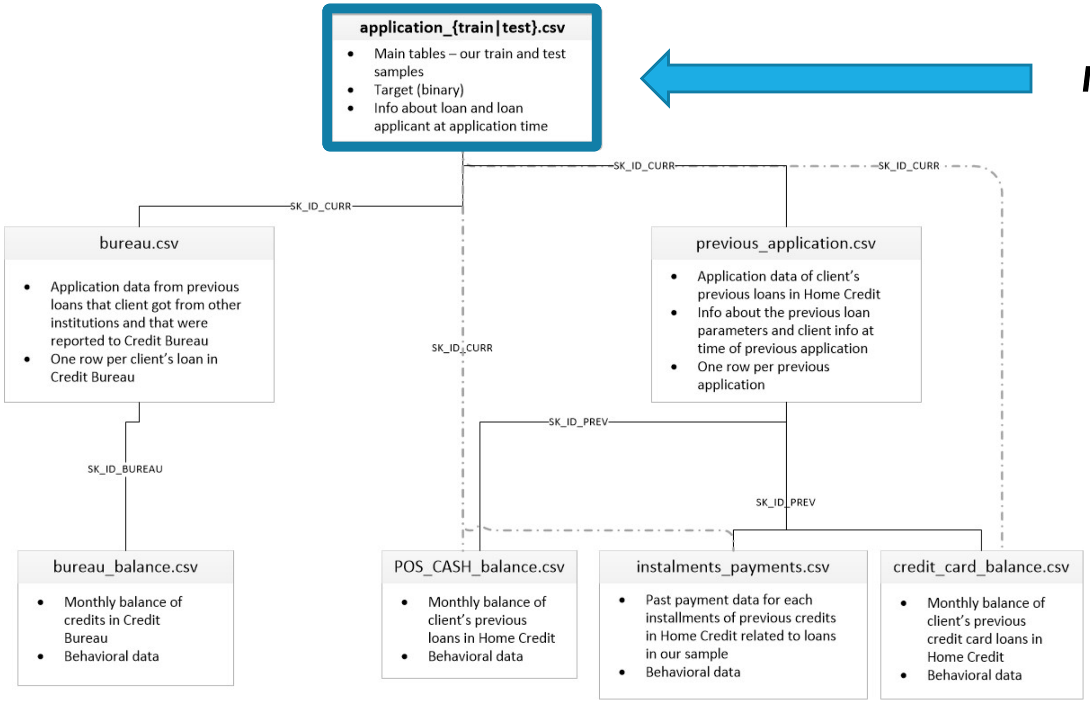
Here’s the link (register for free in order to download the data): https://www.kaggle.com/c/home-credit-default-risk

OUR PLAYGROUND (2/3)

Why is it a good dataset to study ethical implications of AI?

1. It's **rich of information** about the applicants
2. Approving or denying a loan is a **high impact task** as it can be a life changing decision: it's not a ML problem like recommending a movie ☺
3. The data is **anonymized but real** (you can check the company website)
4. The analysis possibilities are **endless**! We'll focus on the core table, but many more information is available

OUR PLAYGROUND (3/3)



MAIN FOCUS

OTHER AVAILABLE DATA

LET'S DIVE IN! (1/6)

Analyse the dataset

Load the data (application_train.csv – the <u>only file</u> we will use) and provide a basic EDA (Exploratory Data Analysis)

#1 Focus on variables that you consider prone to ethical discussions: what are they? Highlight and discuss them

#2 Take a look at the training set: are the interesting variables at #1 related to the target variable? According to their type (binary, continuous, etc.) provide a numerical and/or visual representation

Hint: For instance, take a look at the *gender* variable. Do males and females have the same insolvency ratio, on average? And what about age?

LET'S DIVE IN! (2/6)

Feature importance

Split the dataset in train and test (80%-20%) and then train a simple ML model (you can just use the provided code). Evaluate the *importance* of the different features on the model

#3 Are the features you found in #1 considered relevant by the trained ML model? Yes, no, or better: how much?

Hint: Using the provided model, you can start from the standard built-in *feature importance*[1]. A fancier (and extremely effective) alternative is using SHAP[2] values.

[1] https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.Booster.get_fscore
[2] https://github.com/slundberg/shap

LET'S DIVE IN! (3/6)

Scoring the test set

Let's take a look at the performance of the model on the test set

#4 Ignore the original test target and just take the predictions of the ML model (see provided code). Overall, how do the variables at #1 relate to these predictions? Is it similar to what you saw on #2 and how do you explain it?

Hint: You can exactly rerun the code used on #2. In #2, you analysed some original train features vs. original train target, on #4 you are analysing test features and test predictions provided by the ML model

LET'S DIVE IN! (4/6)

Analyse some test records

Just take some records from the test set (let's say 2 or 3) and look at the predictions of each one

#5 Try and manually alter the sensitive variables (those at #1) and score the new, *altered* records: do the predictions change?

#6 You've analysed the overall behaviour of your ML model on the test set (#4) and even double checked on some individual cases (#5): what's your conclusion? Did the model learn the differences and biases in the original dataset?

LET'S DIVE IN! (5/6)

Remove the sensitive variables and see what happens…

Let's just remove all the variables of some ethical concerns

#7 Retrain the ML model, score the test set and perform the analysis (#3). What's the new AUC performance (with respect to the original one)?

#8 Analyse both the overall behaviour and individual one (#4 to #6). What's going on? Do we still see differences for the average predicition of different groups?

Hint: Should we consider gender, remove the variable but keep track if a record belongs to one group or another (M or F)! This way you can analyse if removing the variable is enough or somehow the differences in the average predictions remain…

LET'S DIVE IN! (6/6)

Understanding deeply the problem and… what's going on

Time to wrap up

#9 Did results at #8 surprise you? Can you explain why just removing the variables wasn't enough?

#10 (difficult) Any ideas on different ways to reduce the bias in this specific problems? What are the difficulties and the tradeoffs we could encounter?

RECAP

#1 to #6 Understand the data and the business problem we're facing: what concerns you from an ethical viewpoint? See how a simple (but not trivial) ML model learns the characteristics and the biases in a real-world case – **MANDATORY**

#7 to #9 Try the simplest way to address your concerns, aiming at a bias-less ML model. Did you solve all the problems? How about the tradeoffs? – **OPTIONAL**

#10 Are there other alternatives to deal with biases in your dataset? (i.e. what AI fairness solutions are all about) – **OPTIONAL (and quite difficult!)**