WFAIS.IF-XG325.z: Explainable Artificial Intelligence

Problem Set 2

Instructor: Dr. Marcin Abram
E-mail: `marcin.abram@alumni.uj.edu.pl`

Deadline: Friday, April 11, 2025 at 11.59 pm CET

As long as the problem set is open, you will be able to send multiple answers (the last submitted answer will matter).

Task (30 point)

You are a new hire in a mid-size company. You have just completed your first task. Today morning, you handed your report in, and now you circle around the office. You feel exhausted and a bit nervous. You don't really know what to expect next...

After lunch, your technical manager approached you. "Our boss wants to see you—we have to discuss the next task for you," he said.

Following your manager, you enter the meeting room. The CEO and the senior developer are already inside.

Your CEO said: "You did an excellent job! We are truly lucky that you have decided to join our firm. Our clients were delighted with the model that you designed for them. The report was very helpful. It highlighted the importance of correctly understanding the model's limitations and the expected performance after it is deployed in the real world.

However, we have another task for you now. We want to start a trial with a major hospital. I want you to prepare a proof-of-concept so we can convince them that a partnership with our firm can benefit them.

You will get historical medical data. I want you to design a model that can classify whether a particular treatment is recommended for the patient or not. Additionally, there are five additional features (denoted as `MeasureA`, `TestB`, `GeneA`, `GeneB`, and `GeneC` in your dataset) that we can use. However, they are really expensive and difficult to collect. I want you to assess how useful they are.

We will meet with the hospital in two weeks. I want a detailed report describing your main findings, on my desk, on Friday, April 11, at 3 pm PDT."

Hint: Your boss called your task "proof-of-concept" but in fact, the nature of that assignment is the same as the last time. You are asked to train a classification model, and you must measure how good that model is. Additionally, you must give recommendations on which features are important to collect. You should look at all variables, but at minimum, you should test the importance of ***`MeasureA`****,* ***`TestB`****,* ***`GeneA`****,* ***`GeneB`****, and* ***`GeneC`****.*

Your Technical Manager said: "This time it really matters, that your model has a good performance. It would be a big deal if we can show that our model makes fewer mistakes than a human doctor. Describe exactly how you tested your model. They are really going to look at that section. Additionally, similar to the last time, the interpretability of the model is very important. You should restrict yourself to logistic regression."

Hint: Remember that accuracy alone is not a good measure. We care about both accuracy and precision. Reports also include false positives and false negatives. To choose the right model, you can use, for example, the AUC score. It is ok (it's even expected) that you will do some feature engineering. You can also try to add regularization to your logistic regression[1] *and test if it helps you or not. To show that the model can be interpreted, you can identify and explain the most important relations between the variables and the expected outcome (e.g., how the probability that the treatment is recommended changes with age or gender)."*

The senior developer took you aside and said: "My task will be to prepare a technical demo based on your work. To do this, I need your code. Remember, that I'm not a Data Scientist like you – so you have to be very careful when you are writing your code. Write comments and try to explain any nontrivial section – so I'm not confused when I read your code."

Hint: In the ideal case, people should be able to take your code, run it, and recreate all your results. In a less-than-ideal case, it should be a demonstration of a "typical run. The code should demonstrate your approach end to end. People should specify the dataset's path, run it, and see the final results.

Learning Objective: You will be able to train and evaluate logistic regression. You will be able to create a technical report. You will be able to communicate your findings with several people who represent different archetypal roles.

Data

First, enter the competition by accepting the terms at `https://www.kaggle.com/t/d296cddb2f144dd4b6d6e469fe9e4bd8`. Then, you will be able

[1] Check the `penalty` parameter in `sklearn` LogisticRegression function. Read more on `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html`

to find the dataset at `https://www.kaggle.com/competitions/uj-xg-325-xai-spring-2025-ps-2/data`. There are two files. The first file, `ps2_available_dataset.csv` should be used to complete this assignment. The second file, `ps2_kaggle_predict.csv` can be used for the (optional; non-obligatory) in-class Kaggle competition.

Report Submission

Your submission should consist of one single pdf file. You should include three separate, self-contained sections:

1. Short, under one page, executive summary for the CEO.
2. A technical discussion (with plots) for the technical manager.
3. A self-contained print-out of the code for the senior developer.

Please, do not print-out all the code you have produced. Try to rather show one, final, self-contained, end-to-end example.

Grading Rules

In order to grade your work, we will role-play the following situation. I will assume that you are a new employee. You were asked to provide a comprehensive technical report that illustrates your findings. We will evaluate each subsequent part from the perspective of the following people.

- Your CEO (she would like to hear the high-level stuff; She will only read the executive summary). *(10 points)*
- Your manager (he would like to see a detailed discussion; he is also very keen to see some plots). *(10 points)*
- A senior developer (they would like to see the code; they might not read the rest of the report at all). *(10 points)*

Your final score is the sum of the scores given by each person. Cumulatively, there are 30 points. Each person (the CEO, the manager, the developer) will use the following grade rubrics.

Grade Component	Meets Expectations	Approaches Expectations	Needs Improvement
Completeness			
Clarity and Support			
Validity			

By “Completeness” I mean that all parts of the question are addressed. By “Clarity” I mean that the text and the code are written in an accessible way. By “Support”, I mean that you provide sufficient evidence to back up your statements. You must cite any source you use (even if you happened to copy or adapt a snippet of code from the internet – you should still treat it as a citation. You should clearly mark how large that snippet is and provide adequate references). In this class, taking code generated by various LLM-based systems is treated the same as copying the code from the internet. It is allowed as long as you are transparent about that (and you cite the source, i.e., you disclose the provider, the name, and the version of the model, e.g., OpenAI’s ChatGPT o1-mini or Anthropic’s Claude 3 Haiku, etc.).

Optional Challenge

We also created an optional challenge for you. There is no additional credits for participating in it[2]. However, we encourage you to try. To try it, go to `https://www.kaggle.com/t/d296cddb2f144dd4b6d6e469fe9e4bd8`. Then, you will be able to find the dataset at `https://www.kaggle.com/competitions/uj-xg-325-xai-spring-2025-ps-2/data`, called `ps2_kaggle_predict.csv` – it is a unique test dataset where I removed the “treatment” column. Train a model (you are not restricted to linear models anymore – you can use whatever you want) and make your predictions. During our lecture on Sunday, April 13, I will present the leaderboard.

Have fun!

Don’t Panic

Don’t panic. I understand that this is a large, open-ended task. I also understand that this might be only your second technical report that you were tasked to write. I am dedicated to helping you do the best work, and while I keep high standards for you, at the same time, I acknowledge that you have limited time and limited resources to complete your task. This report doesn’t have to be perfect to be graded at 100%.

If you don’t know where to start, read the *third* chapter of “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”, 2nd Edition by Aurélien Géron. Check also the *Appendix B. Machine Learning Project Checklist* from that book. And as always, if something is not clear, ask questions using the Pegaz’s forum.

Learning Objective: You will be able to design a machine learning pipeline. You will

[2] Grading students based on the in-class competition is a very bad educational technique! We try to avoid it in this class. However, creating a small competition, and making the participation optional – this could be fun!

be able to create a technical report. You will be able to communicate your findings with several archetypal people.