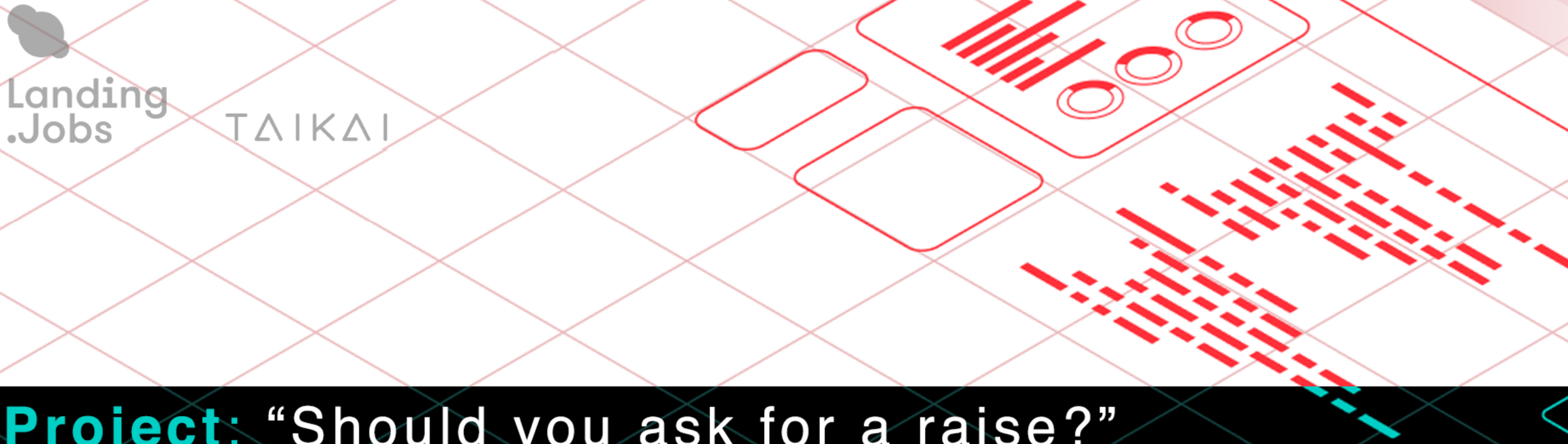
Landing.Jobs Data Challenge



Project: “Should you ask for a raise?”

Author: Wilson Capitão

Hi everyone!

The goal of this project was to develop a simple model able to predict if someone in the tech world, living in Portugal, should earn more than 30k Gross Per Anum, no matter their gender, nationality or age.

I wanted to build a tool that could really help anyone making data driven decisions with the best insight possible.
Asking for a raise is not always the easiest decision. We must know our knowledge, experience and value for the company to understand if we deserve that raise.

Inequalities might bring some uncertanties and lack of confidence for some groups, women for instance. Specilly for these groups, **I hope this tool empowers them to demand fair salaries with no fear.**

Key insights

Landing.Jobs already made an interesting resume and overview about the available data and we also can see what data was collected.
Some points that deserve to be remembered:

- We already know that there is a **big difference between Male and Female wages** from Landing Jobs Report
- The variables "Gender", "Age" and "Nationality" are ethically questionable. All those features are born with us and are part of us. Yes, we can change them, but they shouldn't interfere with our wage. We should be evaluated by our skills. In order to create an ethical model and to empower anyone, no matter their gender, citizenship or age, I've removed those features from the model.
- There are particular skills and characteristics that are prone to higher salaries and the model should rely only on that information

The model

To solve the proposed problem, the data was transformed to be able to be processed by the model. Some categorical variables were labelled and others were one hot encoded - this is the reason for some features in the final model to be specifically about one city or role.

Logistic Regression was chosen because it was the best option to get interpretability and a reasonable accuracy value.
This model was used to get this top 20 list of features, ordered by their importance. "Working Experience" was in almost all experiments and different models the most important feature for prediction.
Almost all features in this top 20 are about the company, the role or where the professional lives. Only "Education Level", "English Level" and "Language Perl" are directly about skills.

After improvements, the final model was able to predict if someone should earn more than 30k€ with 82.62% accuracy.