Natural Language Project

Find the "Chef"

Alameda and Tagus

2025

1 Goals

Simulate a participation in an evaluation forum (max score: 4)

International evaluation forums (ex: CLEF, SemEval, etc.) are competitions in which participants test their systems in specific tasks and in the same test sets. Training sets are given in advance, and, on a certain predefined date, a test set is released. Then, participants have a short period of time to return the output of their systems, which are evaluated and straightfowardly compared with one another, resulting in a final ranking where the state-of-the-art system is acknowledged. We will simulate such an evaluation forum, although the test set (without the expected output) will be released along with the train set.

Develop your critical reasoning, your communication skills, and your creativity (max score: 16)

Write a short paper (max 2 pages) in which you describe the models you created, present the obtained results, discuss the latter, and highlight limitations associated with the given data and your system. You should convince us that you have looked at the data and at the returned outputs (not just at the evaluation scores).

2 Tasks

This project is about associating a chef to a recipe. Your task is:

a) **to build a model**, in Python 3, that maps each recipe to its creator chef. That is, being given a file with a list of N recipes and their features, your system should return another file with N predicted labels, each corresponding to a chef identifier (6 chefs in total);

b) **write a short paper** describing your work.

3 Your Model

You will be given a training set (train.csv), where each line has the following format (notice that there are tabs between the different fields):

chef_id recipe_name data tags steps description ingredients n_ingredients

Example:

5060 ginger chicken chowder 03/02/2003 ['60-minutes-or-less', …, 'taste-mood', 'equipment'] ['melt butter in […] pinch of fresh grated ginger'] mmmmmmm good!! spicy ginger chowder ['butter', […] 'ginger'] 17

For the automatic evaluation of your project, you should run your **best model** on the given test set (test-no-labels.csv). Notice that it has no labels, that is, no "chef_id". The system should return, an output file (that you should name results.txt), each line with:

chef_id

Notice that the line number in which a recipe appears in the test file (test-no-labels.csv) should be the same line number of the corresponding chef_id in results.txt (the automatic evaluation depends on this).

4 Details

- In the paper you should report the results on your own test(s) set(s). However, you don't need to create new datasets: just extract them from the training set.
- To build your model(s), you can use whatever approach you want, including taking advantage of code already available (and we strongly advise you to do so). The only constraints are: a) you should use Python 3; and b) you must identify all the sources you used.
- This project should preferably be done in groups up to 4 students (cooperation). However, groups of a single person are also allowed. If you are looking for a colleague to create a group, please check the sheet we will make available at Fenix (Section Projects).
- Although we will probably not run your code, you should be prepared to do it. If you are unable to replicate your own work, you will have zero in the automatic component.
- Questions should be sent to meic-ln@disciplinas.tecnico.ulisboa.pt (subject: Project).
- We might release FAQs about the project – again, Fenix (Section Projects). Stay tuned!

5 Evaluation

Automatic Evaluation (4 points):

- Accuracy will be the evaluation measure.
- If you beat a weak baseline (Support Vector Classifier, with TF-IDF features extracted from the "description" field) by obtaining an accuracy above 30.0%, you will have 2.0 points.
- If you beat a stronger baseline (Support Vector Classifier, with TF-IDF features extracted from the whole set of fields) by obtaining an accuracy above 43.0%, you will have more 2.0 points.

Short paper Evaluation (16 points):

The short paper should be named NUM.pdf (NUM is the number of the group). It should have a maximum of 2 pages[1], be written in Portuguese or English, and should follow the given template (mandatory). The scores of each section are detailed in the template.

Notice that 3 points will be discounted if any instruction is not followed.

6 Submission

On October 15th, 2025, you should deliver, until 11:59 PM (23h59), via Fenix, a zip file (NOT a rar) containing the project, named after the group number NUM (ex: 3.zip). The zip file should contain:

- the file NUM.pdf with the short paper;
- the project code (not the models) (in .py files or in a .ipynb);
- a file named results.txt with the results from the given test set, that is, a list of the labels returned by your best model when it was applied to the given test set (test_no_labels.txt).

7 Comments/tips

- This is not a B.Sc project; this is a M.Sc project: there is a clearly identified problem that you need to solve in the best possible way, but we do not tell you how to do it.
- Remember what you have learned during the class about methodology: try to do a systematic work. Automatically evaluate your models every time you (try to) improve them.
- Pre-processing applied to the training set should also be applied to the test set.
- Attention to blindly removing stop words.
- Understand that language is too complex to deal with each example individually; also remember that your model will need to be able to generalize.
- There is no 100% accuracy (this is a research problem).
- The dataset is adapted from a "real" dataset from Kaggle[2]. Datasets in NL have errors and are sometimes unbalanced. You will also probably find many labels that you don't agree with. You are probably right, but the datasets will not be changed. Discuss these situations in your short paper.
- Look at the input and output!!!!!!!!!!!!!!!!!! (not just to numbers)

Thank you!

We really hope you **enjoy** the project and have a **good learning experience** with it! ♥

[1] If the report has more than 2 pages, we will only evaluate the first two, even if the first one is a cover page with your numbers and names.

[2] https://www.kaggle.com/