

WideBot – AI assignment

Please find the details of the tasks below and create a Github repo per task.

Use your preferred programming language. We recommend using python.

Please add comments to your code:

Task 1 - Spelling Checker

Using this [dictionary](#), implement a spell checker class that takes this dictionary as input, this class has three main operations:

- Store this dictionary in a suitable data structure.
- Take an input word and return the nearest 4 words if this word is not in the dictionary
- Take an input word and add this word to the dictionary

For each operation specify the time and space complexity

Note: You could assume that the nearest 4 words from a word are the 2 words before and after this word in lexicographic order if they exist.

Task 2 - EDA

Using this [dataset](#), you are supposed to prepare an Exploratory Data Analysis (EDA) report in PDF format that shows at least 3 different insights about this data (number of examples per class, top frequent n-grams generally and per class, lengths of examples in words and letters,) . Feel free to show your insights in a good format (description, tables, charts, ...).

Note that you can prepare the PDF manually after generating the analytics in any other format (CSV, xls, doc), ...).

Note: Use stories data only (not comments).

Task 3 - Classification

Using the same data in task 2, build a machine learning based classifier and show its performance on the test-set. Show precision, recall, f-score, accuracy for each class and for the whole test data. Please describe briefly the meaning of each result and metric for this specific task. Also, please write some enhancements that you may think about to achieve better results.

Note: add a readme file that describes the whole training process in briefed points. Note: Use stories data only (not comments). Also, take the last 20% of each file as a test-set.

All the best 😊