

Text Classification with Python Project

Project Deliverables

You will be required to complete the following deliverable.

- A python notebook with your solution.

Instructions

Background Information

You work as a Computational Linguist for a Global firm, collaborating with Engineers and Researchers in Assistant and Research & Machine Intelligence to develop language understanding models that improve our ability to understand and generate natural language.

Problem Statement

You have been tasked to build a text language classification system that classifies a given input as written in English or in Dutch.

You can use the following Guiding Template [[Link](#)].

Dataset

You have been provided with the following dataset that contains over 1000 examples of around 15 words segments along with a class specifying if the words in the example are written in English or Dutch. Each record begins with a two letter class code describing the class to which the following record belongs i.e. "en" - stands for English and "nl" - stands for Dutch.

All the words of a record belong only to one class. There is no record that contains some words belonging to "en" and others belonging to "nl".

- **Dataset URL (CSV):** <http://bit.ly/EnglishNDutchDs>

Acknowledgements

- <https://bit.ly/3ISLucy>