

Natural Language Processing (CS5803): Assignment-2

Text Classification

Assigned on: 9th Feb, 2025
Deadline: 22nd Feb, 2025 11:59 P.M.

Problem Statement

The objective of this assignment is to independently develop a classification model for each of the two provided datasets. You will experiment with four different algorithms (excluding transformers) and analyze their performance.

Dataset:

One dataset will be provided with this assignment, and a second dataset will be made available next week.

Dataset 1 (Provided with this assignment):

Contains titles, abstracts, and topic labels of research articles. Your task is to predict the topic based on the title and abstract.

Dataset 2 (will be provided shortly)

Tasks:

- Implement and experiment with four different machine learning algorithms (excluding transformer-based models).
- Train and evaluate the models using appropriate metrics.

- Compare the performance of these models and analyze their strengths and weaknesses.

Report Submission:

Your submission should include a well-structured report covering the following:

- Methodology: Explain your approach, preprocessing steps, and modeling choices.
- Models Used: Describe the four selected algorithms and their implementation details.
- Results & Analysis: Present evaluation metrics, compare model performances, and provide insights based on the results.

Submission Guidelines

- Submit a zip file named `<group member RollNumbers>.zip` that contains
 - Your Jupyter Notebooks (.ipynb) file
 - report (.pdf) file
- Ensure that your code is well-commented and easy to understand.

If you have any questions, contact the TA of the course at least 2 days before the deadline.

Happy Coding!