

Time

- 7 days to develop
- Recommended time for all tasks: up to 8 hours

Results delivery

- Link to code repository with the solution (preferably Jupyter notebook).
- Classification results description.
- We will assess whole workflow, not only the performance of the model.

Dataset

- Based on 'IMDB dataset' (attached in zip file)

Task

- Classification
 - Prepare dataset to be usable for modeling.
 - Provide exploratory data analysis part (EDA).
 - Build sentiment classifiers and obtain predictions from at least 2 different models.
 - Combine predictions using one of ensembling techniques.
 - Provide performance summary on the train, val & test sets.
 - Provide explainable AI part that explains what factors drive predictions, what are the most influential features, etc.

- Anomaly detection

Anomaly detection on 'review' (ignore 'sentiment' column)

- Prepare dataset to be usable for anomaly detection task.
- Create a model to detect outliers/anomalies in this dataset.
- Describe what are the anomalous rows and why were they detected as anomaly (which factors influenced)

- Result summary
 - Create slides (up to 7 slides).
 - Explain technical approach in Classification task, briefly describe techniques and packages used.
 - If unable to solve both objectives, explain what the next steps would be to do so
- Notes
 - Please follow PEP8 and general programming best practices.
 - Choice of Python packages and algorithms is up to you.