

Advanced DS for Innovation

Assignment 2

The Brief:

You have been tasked to work on a project to deploy a Machine Learning model into production. You will train a machine learning model (using sklearn) that will accurately predict a type of beer based on some users' rating criterias such as appearance, aroma, palate or taste. You will also need to build a web app and deploy it online (using Heroku) in order to serve your model for real time predictions.

Deliverables:

- Zip file containing project code, model artifacts, instructions or any relevant documents
- Final report describing the machine learning model and its hyperparameters , performance achieved, API structure, instructions for running predictions. The report should not exceed 1500 words. Don't forget to include links to your github repository and API url.
- Deployed model to Heroku

Note: Your best model needs to be accessible online on Heroku. A series of test will be performed on it and will be part of the marking

Here are the list of expected accessible API endpoints:

- '/' (GET): Displaying a brief description of the project objectives, list of endpoints, expected input parameters and output format of the model, link to the Github repo related to this project
- '/health/' (GET): Returning status code 200 with a string with a welcome message of your choice
- '/beer/type/' (GET): Returning prediction for a single input only
- '/beers/type/' (GET): Returning predictions for a multiple inputs

Submission:

All assignments need to be submitted before the due date on Canvas. Penalties will be applied for late submission.

Assessment Criteria:

- Quality of code, repository, robustness of model deployed
- Justification of architecture selected, data transformation performed, hyperparameters selected and accuracy of results with evidence supporting claims
- Quality of findings and recommendations

- Clarity and quality of visualizations and written report

Dataset:

This dataset contains around 1.5 Millions reviews of Beer from BeerAdvocates.

Link to dataset:

https://drive.google.com/file/d/1IIOsD9n_HpdqZ5Ha1eEtIPwROqoEb223/view?usp=sharing

Data Dictionary:

Column	Description	API Expected Parameter
brewery_id	Identifier of brewery	No
brewery_name	Name of brewery	Yes
review_time	Timestamp of review	No
review_overall	Overall score given by reviewer	No
review_aroma	Score given by reviewer regarding beer aroma	Yes
review_appearance	Score given by reviewer regarding beer appearance	Yes
review_username	Profile name of reviewer	No
review_palate	Score given by reviewer regarding beer palate	Yes
review_taste	Score given by reviewer regarding beer taste	Yes

beer_style (target)	Type of beer	No
beer_name	Name of beer	No
beer_abv	Alcohol by volume measure	Yes
beer_beerid	Identifier of beer	No