

Assignment for MLOps 2025

Submission deadline: Friday 21.3.2025 at 23:59

Submit by sending your Github Repository to me (Primoz) on Teams

Penguins of Madagascar



About the Assignment

We are planning to go to New York and would like to find Skipper, Private, Rico, and Kowalski! We don't know much about them, except their species: Adelie.

Adelie penguins live on the coast of Antarctica and their conservation status is Near Threatened. We would like to find these 4 penguins and bring them home to reproduce!

Every day at 7AM we will get data about a new penguin spotted in the streets of NY. You need to make a classifier to tell its Species. Given data about the penguin will be: **'bill_length_mm', 'bill_depth_mm', 'flipper_length_mm', 'body_mass_g'**. You will get data at this API: http://130.225.39.127:8000/new_penguin/

Task 1

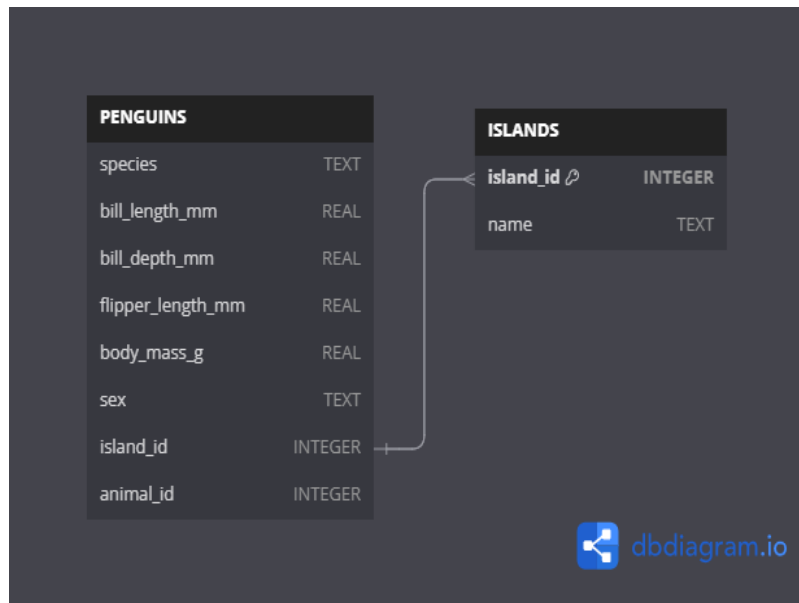
Create new (public) GitHub Repository, which contains Readme file. Readme file should contain all technical details about the submission.

Task 2

Download dataset (look at the snippet), and transform data in to database (look at the proposed schema).

Save database in your repository! Have a separate file for this step!

```
# Load the penguins dataset. It requires package 'seaborn'
penguins = sns.load_dataset("penguins").dropna()
penguins.info()
```



Task 3

Perform relevant feature selection and provide your decision making process. Read data from SQL database.

Split dataset to training and testing, perform transformation (if needed) and train any suitable classification model.

Save the model in the Repository! Have a separate file for this step!

Task 4

Every morning at 7:30 AM, fetch data from API and give prediction on Github Pages.

Set GitHub Action to fetch new datapoint, make and save prediction!

Keep your Repository organized. Save data in separate folder. In Readme.md, describe the repository and explain its functionality, and provide used technical things.