First project:

You will have to form a group of 3 to 4 students.

Choose a dataset from the UC Irvine Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php) with at least 5000 instances and 20 attributes for classification or regression. Compare how the different approaches seen in class perform on this dataset to predict accurately the classes or the values of the unlabeled data. You should determine what are the best hyper-parameters for each approach you are using. You could use any Python libraries.

<https://archive.ics.uci.edu/dataset/494/wave+energy+converters>

**Wave Energy Converters**

This data set consists of positions and absorbed power outputs of wave energy converters (WECs) in four real wave scenarios from the southern coast of Australia.

A close-up of a computer code

Description automatically generated

**Attribute Information**

Additional Information

Attribute: Attribute Range

1. WECs position {X1, X2, â€¦, X16; Y1, Y2,â€¦, Y16} continuous from 0 to 566 (m).

2. WECs absorbed power: {P1, P2, â€¦, P16}

3. Total power output of the farm: Powerall

Data Source: The data is collected from real wave scenarios at four locations along the southern coast of Australia: Sydney, Adelaide, Perth, and Tasmania.

Number of WECs: There are 16 wave energy converter (WEC) locations considered in the study.

Size-Constrained Environment: These 16 WECs are optimized within a size-constrained environment. This means that there are limitations on the physical size or placement of the WECs.

papers:

<https://arxiv.org/ftp/arxiv/papers/2011/2011.13130.pdf>

<https://www.youtube.com/watch?v=4yHHF9goNqE>

Wave Farm

<https://www.youtube.com/watch?v=-XWbVMtNnaw>

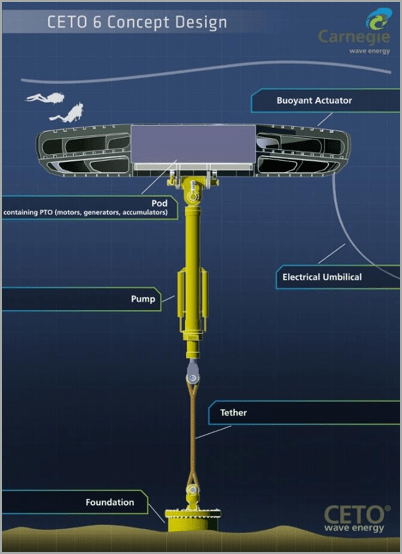
A group of yellow buoys floating in the water

Description automatically generated

A computer screen with graphics

Description automatically generated

CETO



Pages: 5 - 8 pages

- include a presentation of the research questions

- the chosen methods to tackle them

- a presentation of the results and discussion and a conclusion/future work

- attach a description of the participation of each student to the project

Wave Energy Converters (WECs) convert wave power into electricity.

* 6 WECs locations are placed and optimized in a size-constrained environment.
* The data set consists of 48 attributes, out of which the first 32 attributes are the positions (latitude and longitude) of the 16 WEC, continuous from 0 to 566 (m).
* The next 16 attributes are WECs absorbed power.
* The last attribute is the total power outputs from the farms.
* The WECs positions are used to predict the total power outputs.
* A total of 72000 rows of data is present in each dataset.
* Each row consists of the data from the wave farm by changing the positions of the 16 converters and their respective absorbed power is noted.

[1] - https://arxiv.org/ftp/arxiv/papers/2011/2011.13130.pdf

**The first data frames long & lat. is plotted like this**  
A screen shot of a graph

Description automatically generated

**First Project: Machine Learning Analysis**

**Research Questions:** The project will address research questions related to the optimization of Wave Energy Converters (WECs) in a size-constrained environment.

**Methods:** The chosen methods will involve utilizing machine learning algorithms to predict total power outputs based on the positions and absorbed power of WECs.

**Abstract**

This document outlines the objectives, methodologies, and deliverables for the first project, focusing on machine learning analysis. The project aims to explore and compare various machine learning approaches to predict classes or values for unlabeled data in a dataset from the UC Irvine Machine Learning Repository. It involves forming a group of 3 to 4 students and utilizing Python programming with relevant libraries. The project will follow a structured process, including dataset selection, preprocessing, model selection, training, evaluation, hyper-parameter tuning, result comparison, and documentation.

1. **Introduction**

The first project involves conducting a machine learning analysis on a dataset from the UC Irvine Machine Learning Repository. Key objectives include:

1. Forming a group of 3 to 4 students.

2. Selecting a dataset with a minimum of 5000 instances and at least 20 attributes suitable for classification or regression.

3. Comparing different machine learning approaches to predict classes or values for unlabeled data.

4. Determining the optimal hyper-parameters for each approach.

5. Utilizing Python and relevant libraries for the analysis.

1. **Dataset Selection and Exploration**

- The group will choose a suitable dataset from the UC Irvine Machine Learning Repository.

- A thorough exploration of the dataset will be performed to understand its structure, features, and characteristics.

1. **Data Preprocessing**

- Data preprocessing tasks will include handling missing values, encoding categorical variables, and scaling data for analysis.

1. **Model Selection**

- Multiple machine learning algorithms suitable for regression will be selected.

1. **Model Training and Evaluation**

- The selected models will be trained on the dataset.

- Evaluation of each model's performance will be conducted using appropriate metrics, such as accuracy, F1-score, and mean squared error.

1. **Hyper-Parameter Tuning**

- Identify the best hyper-parameters for each model to optimize their performance.

1. **Comparison of Results**

- Results and performance metrics of different models will be compared.

- The model(s) achieving the best results will be determined.

1. **Conclusion and Findings**

- A summary of findings from the analysis will be provided.

- Implications of the results will be discussed.

1. **Description of Student Participation**

-Attach a description of the participation of each student in the project.

**References**

[UC Irvine Machine Learning Repository] (https://archive.ics.uci.edu/ml/index.php)