Assignment - Machine Learning [Major]

Grading			
Event	Decoding Skills	Number of question not attempted	Overall Output
Assignment	0- If the learner does not submits the assignment or if he tried to attempt it but the applied hypothesis is wrong or showing an error. 10- If the learner clearly decodes the given data set or questions by performing the tasks defined in the question	0 - If the learner does not solves any questions or solves less then 40% of the assignment correctly. 5 - If the learner successfully solves between 40-80% of the given questions. 10- If the learner solves 80-100% of the questions correctly	0-If the output presented is completely wrong. 5- If the given output is partially correct along with incorrect presentation. 10- If all the answers are attempted correctly along with presentation skills

Use the Oil Spill Dataset and solve the following question by using the dataset, to download the dataset click on the dataset name.

About Dataset

The dataset was developed by starting with satellite images of the ocean, some of which contain an oil spill and some that do not.
Images were split into sections and processed using computer vision algorithms to provide a vector of features to describe the contents of the image section or patch.
The task is, given a vector that describes the contents of a patch of a satellite image, then predicts whether the patch contains an oil spill or not, e.g. from the illegal or accidental dumping of oil in the ocean.

There are two classes and the goal is to distinguish between spill and non-spill using the features for a given ocean patch.

- **Non-Spill:** negative case, or majority class.

- **Oil Spill:** positive case, or minority class.

There are a total of 50 Columns in the Dataset , the output column is named as **target.**

- Download the Oil Spill Dataset and perform Data cleaning and Data Pre-Processing if Necessary.
- Use the various methods such as Handling null values, One-Hot Encoding, Imputation and Scaling of Data Pre-Processing where necessary.
- Derive some insights from the dataset.
- Apply various Machine Learning techniques to predict the output in **target** column, make use of Bagging and Ensemble as required and find the best model by evaluating the model using Model evaluation techniques.
- Save the best model and Load the model
- Take the original data set and make another dataset by randomly picking 20 data points from the oil spill dataset and apply the saved model on the same .

Note:- For submission, upload the code file (notebook) and the randomly generated dataset on your drive and share the drive link of the files in a word document.

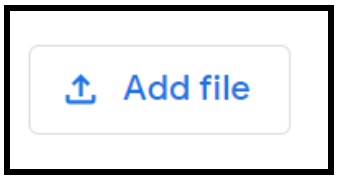
Process for Submission

Please upload your assignment files via this Google-form before the deadline (19th June, 2023 11:59 pm). Upload the assignment file name as **Assignment - Machine Learning [Major] by <your-name>**

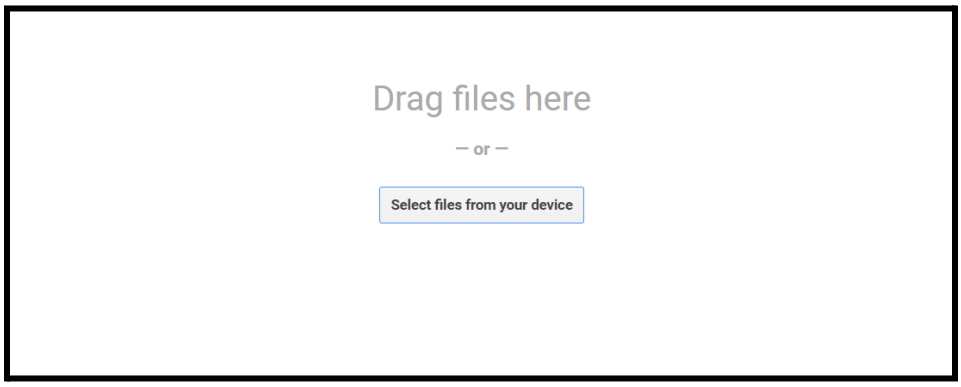
CLICK HERE to Access the **Google-form**.

Steps to submit the assignment:

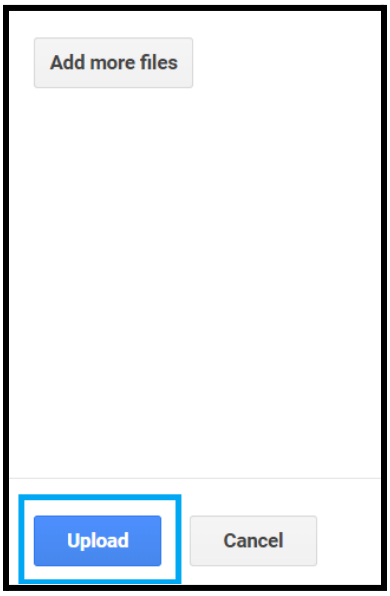
1. Open the Google Form.
2. Enter your email id, name, assignment name.
3. And upload the assignment file by clicking the Add file button.



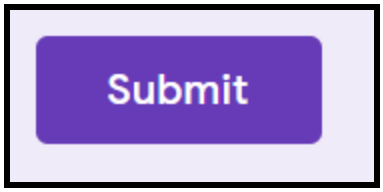
4. You can upload the file by clicking Select file from your device or by dragging the file and dropping it in the window opened.



5. Select your assignment file and click the Open button.
6. After selecting the file click on the Upload button.



7. Once the file is uploaded, click on the Submit button.


Submit