

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**  
**K. K. BIRLA GOA CAMPUS**

First Semester 2023-24  
Machine Learning (BITS F464)  
Assignment

A. Given the following dataset, build a machine learning model for binary classification.

**Problem Statement:** Build a **machine learning model** to detect potential Higgs Boson signal from one of the background processes that mimics it.

**Dataset Description:**

The features present in this dataset are kinematic properties measured by the detectors and functions of these measured properties.

The dataset consists of a single binary target which indicates if a Higgs Boson signal is detected or not.

**B. General Instructions:**

1. Create a Kaggle account (if you don't have one) and join the [competition](#).
2. **Change the Team name to your id: 201XXXXXXXXXG.**
3. Download the train and test data files. You are given a train dataset (0 means background noise and 1 means boson detection) and test dataset with 30 attributes.
4. You need to build a model(s) that can assign given tuples into one of the two predefined classes.
5. You are free to use any external library of your choice.
6. You need to upload a zip file in the format of the sample submission file (given on kaggle) to submit your solution (same as how we have done in previous labs).
7. Your submission will be evaluated on 60% of the dataset to give your result on the public leaderboard. At the end of the competition, your selected submission (You can select maximum **one** submission) will be evaluated on the remaining 40% of the data and will be shown on the private leaderboard. However, the final score will be based on your performance on 100% data. The evaluation metric is **accuracy**.
8. You are given **7 days** with **5 submissions** on each day. You are requested to start working on your model immediately to avoid pressure at the last moment. No requests will be entertained to increase the submission deadline or to increase the number of submissions allowed per day.
9. Select **1 submission** to be used to calculate your standing on the private leaderboard. If no submission is selected, Kaggle will automatically select 1 which might not be the one you intended.
10. We recommend you use colab for training your models.

**C. Answer the following questions in Report:**

1. Discuss your insights about the data.
2. Discuss any preprocessing and feature engineering steps you might have employed.
3. Describe the ML model you have used and the reasoning behind your choice.
4. Discuss evaluation metric(s) you have employed to measure the performance.
5. Discuss any other approaches that you would want to try out in the future.

## D. Final Assignment Submission Instructions:

Same “**FINAL CODE**” should be submitted in both **KAGGLE** and **QUANTA**. In case of a mismatch, **zero marks** will be awarded for this assignment.

### 1. Assignment Submission Format: (**KAGGLE**)

- a. The result obtained on running your uploaded kernel must match your score on the public leaderboard. In case of a mismatch, **zero marks** will be awarded for this assignment.

### 2. Assignment Submission Format: (**QUANTA**)

A zip file (201XXXXXXXXXG.zip) consisting of the following: (each file should be named as “**ID\_\*\*\*\*\***”)

- a. (**ID\_psc**) Portable source code:
  - i. It is advisable to use a single Python notebook for cleaner code and submissions.
  - ii. You can use colab for your experiments but must download the notebook as well as make a requirements.txt file for the same, Colab links would not be accepted.
  - iii. Must contain **ALL** external packages/libraries in a *requirements.txt* file.
  - iv. Path for any required file(s) should not be local to your machine
  - v. Instructor should be able to run your code after direct download.
- b. (**ID\_psc\_pdf**) Final portable source code (pdf format).
  - i. Provide a PDF copy of your final source code.
- c. (**ID\_report**) Report in PDF format (max 2 pages. 11pt. Times New Roman.)
  - i. Answers to the questions stated above must be part of your report.
  - ii. Can include insights, inferences, results, and conclusions that are drawn from the assignment.
  - iii. Proper references to the source code and figures.
- d. (**ID\_README.md**) file with instructions on how to run the code (this can be replaced with a text block in the notebook as well).

## E. Assignment Submission Policy:

- Submission accepted through (**QUANTA + KAGGLE**) only.
- No assignment will be accepted by **email after the deadline**.

## F. Plagiarism:

Plagiarism will be checked for every submission.

- The rule is very simple
- If (**Plagiarism % from Turnitin Report**) > 30
  - o Will be awarded “**Component Maximum Marks \* -1**”