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

First Semester 2023-24
BITS F464: Machine Learning
Random Forest

Given the following dataset, build a **Random Forest** model **only** that can predict the **target** attribute given the other attributes:

1. General Instructions:

1. Create a Kaggle account (if you do not have one) with your BITS ID and join the competition.
2. Change the Team name to your ID: 201XXXXXXXXG.
3. In Kaggle, you have access to 3 files **train.csv**, **val.csv** and **test.csv**. These will be uploaded to Quanta as well.
4. For this evaluation, you will be building on work done in the previous lab.
5. You need to build a **Random Forest** model **only** that can predict the "target" feature of the dataset. Use **train.csv** to train your model, verify your predictions on **val.csv** and make your predictions on **test.csv**
6. You will be given a starter notebook called **eval.ipynb** (uploaded on Quanta). Follow the instructions in the notebook.
7. **DO NOT** change the name of this notebook.
8. After you have obtained your predictions in the format specified in the notebook, run the last code cell. This creates a zip file containing
   a. your notebook
   b. submission.csv
9. Upload this same zip file to **both** Kaggle and Quanta.
10. **DO NOT** upload the submission.csv file on your own to kaggle.
11. To upload your zip to kaggle, click on **submit predictions**. This should then give you an option to upload your zip file.
12. Your submissions will be evaluated on 50% of the dataset to give your results on the public leaderboard. At the end of the competition, your selected submissions (You can select a maximum of **one best** submission) will be evaluated on the remaining 50% 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 F1 Score = (2 * precision * recall / (precision + recall)).
13. You are free to use any resources on the internet/previous labs and tutorials for this assignment. **Any form of collaboration is strictly prohibited**.
14. Use of any external libraries **except the ones listed below is strictly forbidden**:
   a. NumPy
   b. Pandas
   c. Matplotlib
   d. tqdm (If you want to use)
15. Examples of libraries you are not allowed to use include **(Not exhaustive list)**.
   a. sklearn
   b. scipy
   c. tensorflow/pytorch
16. Kaggle Competition Link (will be enabled only for the first 5 minutes of the competition).

2. Final Lab Submission Instructions:

The same "final code" should be submitted in Kaggle and Quanta.

Assignment Submission Format: (KAGGLE and Quanta)
- Submit the **same** zip file on both Kaggle and Quanta.
- If the 2 files submitted are different, your submission will be considered null and void.
- The results obtained on running the notebook should match the results uploaded to Kaggle.
- To avoid the issue of randomness, we have seeded the code.

3. Assignment Submission Policy:

- Submission is accepted through **Kaggle and Quanta only.**
- No assignment will be accepted by **email** or any other form **after the deadline**.

4. Plagiarism Policy:

- **Discussing assignment with TA/peer come under plagiarism and will be awarded (-20 Marks)**
- Plagiarism will be checked **AUTOMATICALLY** for every submission.
- In cases of plagiarism, **"Component Maximum Marks * -1"** will be awarded.