**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**K. K. BIRLA GOA CAMPUS**

First Semester 2023-24

BITS F464: Machine Learning

**Logistic Regression**

Given the following dataset, build a model **using logistic regression only** that can predict the **target** attribute given the other attributes:

1. **General Instructions:**
2. Create a Kaggle account (if you do not have one) with your BITS id and join the competition.
3. Change the Team name to your id: 201XXXXXXXXXG.
4. In Kaggle, you have access to 2 files **train.csv** and **test.csv**. These will be uploaded to quanta as well.
5. You need to build a model(s) using **Logistic Regression only** that can predict the “target” feature of the dataset. Use **train.csv** to train your model 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
   1. your notebook
   2. 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 73% of the dataset to give your results on the public leaderboard. At the end of the competition, your selected submissions (You can select maximum **one best** submission) will be evaluated on rest 27% 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 = (correct predictions / total predictions)'.
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**:
    1. NumPy
    2. Pandas
    3. Matplotlib
15. Examples of libraries you are not allowed to use include **(Not exhaustive list)**.
    1. sklearn
    2. scipy
    3. tensorflow/pytorch
16. [Kaggle Competition Link](https://www.kaggle.com/t/643399e993cd47dcac2d97a1fd0a0101) (will be enabled only for the first 5 minutes of the competition).

<https://www.kaggle.com/competitions/eval-lab-1>

1. **Final Lab Submission Instructions:**

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 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 with randomness, we have seeded the code.

1. **Assignment Submission Policy:**

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

1. **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.