**CSCI 555-20 Data mining And Machine Learning**

The Toxicity Prediction Challenge - README FILE

**Team Name**: Bit and Byte

**Team Members**: Shreya Dandavate - x2020gdv, Dhwani Patel - x2020fvk

**Preloaded Files in Zip**

Provided zip file contains all the necessary below mentioned files.

|  |  |
| --- | --- |
| **File Description** | **File Name** |
| Source Code | finalcode.ipynb |
| Training Dataset | train.csv |
| Testing Dataset | test.csv |
| Feature Matrix | femat.csv |
| Project Report | FinalReport.pdf |

**Files Generated After Successful Compilation**

|  |  |
| --- | --- |
| **File Description** | **File Name** |
| Kaggle Submission File | Bit\_and\_Byte\_submission.csv |

**Downloading Packages:**

All the necessary packages will be imported during the execution of the ‘finalcode.ipynb’, therefore, no need to download any package explicitly

**Instructions on how to run the code:**

**Step-1:**

Open Google Collab → File (Top Left) → Upload NoteBook → Select the file → Press OK

**Step-2:**

Left vertical navigation bar → Select ‘Files’→ Select ‘Upload to session storage’  → Select Three files ‘feamat.csv’, ‘train.csv’, ‘test.csv’ and wait for them to upload → Go to Runtime in navigation Bar → Run all

**Software and Hardware Configuration Used**

|  |  |
| --- | --- |
| **Operating System** | Windows 10 |
| **Editor** | Google Colab |
| **Packages** | Pandas, Sklearn, Numpy, Collections, Imblearn, Xgboost, Time |

·