**Toxic Language Detection Using Online Game Chat Data Set**/

The main purpose of this project is to detect toxic players using online chat data. League of Legends Tribunal Chatlogs was used as data. You can access the dataset from the link below.

<https://www.kaggle.com/datasets/simshengxue/league-of-legends-tribunal-chatlogs?resource=download&sort=votes>

\*\*About Dataset \*\*

Context This dataset is scraped from the League of Legends tribunal cases before it was taken down. The cases.zip file is the original data scraped from Tribunal, containing separate .json files. 1 file is for 1 tribunal case. (Around 10,000 cases) The input-data.json file is the combined file of all the .json files above. The chatlogs.csv is the table format of all the .json files above. (1 million+ chat messages)

Content Offender refers to the single player who is reported in the tribunal case. Each row is a single messsage sent by a player.

* message: The content of the message
* associationtooffender: Same team or enemy team as the offender
* time casetotalreports: Number of reports before this case is brought up to Tribunal
* alliedreportcount: Number of reports of allies in same team as offender
* enemyreportcount: Number of report of enemies in opposing team as offender
* mostcommonreportreason: Most common report reason of the 5 available
* chatlogid: The unique chatlog identification number
* champion\_name: The champion name of the player who sent the message.

During this project, machine learning algorithms like LR, RF, LGBM were used. Model scores are shared below the models.