1) Provide a Background/Overview of your proposed research area. Conclude with the problem you are trying to solve in this area. (Max 250 Words)

The area chosen to explore is to predict the run scored in a cricket match for each ball until the end. After going through some research papers that are related to the cricket domain, predicting for each ball seems to be a plausible statement that can be solved by Machine Learning. In detail, trying to predict what type run the batsman has scored for that specific ball, It may vary between 0 to 6 or wicket. Hence it is a multi-class classification problem, assuming there’s no class imbalance for all classes. Proposed algorithm to be used for this problem are, RandomForest Classifier, Logistic Regression, KNN Classifier, Ada-Boost Classifier, XGBoostRF and CatBoost Classifier, tuning them using GridSearchCV and RandomizedSearchCV. Since the approximate size of the dataset (combining all the T20 matches up to date) is around 450,000 rows and 25 columns, the training period required may be long, hence using CatBoost and XGBoostRF to access GPU memory and achieve parallel processing.

2) Proposed Dissertation Title

Ball by ball run prediction using Machine Learning

3) Research Aims & Objectives (Research Question(s))- Please provide your initial Aims/Objectives or research question(s) to help us understand your proposed research further.

Research Question 1 : Accuracy of the model after training.

Research Question 2 : Model performance comparison.

Research Question 3 : Predicting 2023 year matches, runs scored per ball after training with previous years data.

Research Question 4: Probability estimation or prediction of which team will win the cup for the year 2023.