Question 1: How many runs of experiments were executed and what measures were taken to have consistent reproducibility of the results?

Answer: This research started with objective of fantasy points estimation, and we did many rounds of experiments with different features and then considering the drawbacks of it which I have mentioned in the report, further research was continued with forecasting the fantasy points and in that experiment was done with different window sizes and hyper parameter tuning was performed with different set of values and finally the best case was considered. I cannot exactly tell the number of experiments which i executed but it was carried throughout the semester with a guidance from the guide on every week. The research is done for three players Virat Kohli, Rohit Sharma, and MS Dhoni. There are three cases (code) which are traditional ML models such as Linear Regression, SVR (Polynomial and Linear Kernel) and three deep learning models. So, for each player there are a total of 6 cases have been performed. For reproducibility, the code uses a fixed random state in the train_test_split function and have used fixed window size so that enhanced data remains same, ensuring consistent data splitting in each run. The systematic approach to data preprocessing, model building, and evaluation also contributes to reproducibility.

Question 2: How is this study motivated and stems from existing literature? For instance, there is no reference to existing literature in the importance section on Page 2.

Answer: From the childhood I have always been amazed by the sports and as I grew up I was able to follow it closely especially cricket and the strategies they used to plan made me more interested in that field. Indian Premier League (IPL) made wonders to Indian cricket and since so much is invested in this league, the role of analytics came more into the light as there will is a analytical team for each franchise. The way they study the historical data, find the patterns, make strategies, predictions motivated me to learn more about those things and hence I picked up this for my research.

Question 3: What is the difference between 'estimation' and 'prediction'? Why is it relevant to estimate fantasy points from a practical perspective in cricket? What is the motivation or relevance for this line of experimentation?

Answer: Estimation mainly involves determining or calculating the value of something based on available data and in our case, we are predicting fantasy points based on their batting statistics and one more thing is we consider current match data as well for the estimation, while prediction refers to forecasting future outcomes based on just historical data and patterns in the data. Estimating fantasy points is relevant in cricket for understanding current player performance, crucial for immediate decision-making in fantasy leagues as the stakes are very high in IPL. Prediction is used for anticipating future performances, helping in long-term strategy and player selection. The research is motivated by the practical need to make informed decisions in fantasy sports and team strategies, which has a significant following and impact in the sports industry.

Question 4: Why is bidirectional LSTM a relevant model for such time series forecasting task? Compare its advantages and disadvantages against GRU and LSTM while considering the example of forecasting runs scored by a batsman.

Answer: Bidirectional LSTM processes data in both forward and backward directions, capturing dependencies and patterns which may be missed in a unidirectional approach.

Advantages: Better at understanding context and temporal relationships in sequences, which is very much required in sports where past and recent performances can influence future outcomes.

Disadvantages: More complex and computationally intensive compared to LSTM and GRU. Compared to LSTM and GRU, Bidirectional LSTM looks at both the past and recent performances, making it better at catching the details of a player's form over time compared to other methods like LSTM and GRU.

Question 5: In Table 2, the performance ranking is completely reversed for the three models between training and test datasets. On training data, LSTM is better than Bi-LSTM, which is better than GRU; whereas on test data, GRU is better than Bi-LSTM, which is better than LSTM. Why is it so? Explain.

Answer: Yes, as per the table 2, the performance is reversed for training and testing data. The performance seems to be good for training data and not so good for test data and it can be because of overfitting of the dataset was small and it is not able to perform the same with unseen test data. In test data GRU seems to have better result than LSTM and BiLSTM, but we can not really state so because the difference between BilSTM and GRU is very less, and it also because of the simple structure of GRU. We tried to address this over fitting issue by using rolling window approach in which we tried to make the dataset large, and we even tried hyper parameter tuning and still we could not improve much, may be more data would help to address this issue.