WEEKLY REPORT-2

1. Introduction:

The objective of the project is to assess the influence of cognitive states on athletic readiness, specifically predicting RSImod. This is a regression problem where the goal is to predict RSImod using the other variables as input features.

2. Progress Made:

a) Dataset Acquisition and Exploratory Data Analysis (EDA): The dataset was obtained from the faculty, and we have started performing exploratory data analysis (EDA). This process involves understanding the acronyms of the input variables and examining their qualitative relationships with the target variable (RSImod). This step is essential for gaining a clearer understanding of the regression problem and establishing the correlation between various variables.

b) Data Preprocessing: Efforts are underway to clean the dataset by addressing missing values and handling null entries. We are working on appropriately filling missing values by identifying whether variables are numerical or categorical, ensuring proper data imputation.

c) Literature Review: A thorough review of relevant literature has been conducted to match the existing work with our research problem. From the literature, we identified that similar regression problems, where the goal is to predict Reactive Strength Index (RSI), have been handled using Decision Tree Regressor or Random Forest Regressor, enhanced by XGBoost. Based on this, we plan to adopt similar algorithms for our problem.

3. Key Terminologies:

a) Cognitive Assessment: Refers to the evaluation of mental functions, such as memory, attention, and problem-solving abilities.

b) RSImod: A measure of explosive strength, often calculated using jump height and time to takeoff.

c) XAI (Explainable Artificial Intelligence): Refers to techniques that enhance the transparency and interpretability of machine learning models and predictions.

4. Challenges Encountered:

a) Preprocessing Challenges: The preprocessing phase, particularly data cleaning, has posed significant challenges. Filling missing values in the dataset is difficult, as it requires distinguishing between numerical and categorical variables. This process must be done meticulously to ensure the dataset is complete and accurate.

b) Limited Resources: There is a limited amount of literature directly applicable to our specific problem, which has constrained the breadth of our research. However, the available literature has provided substantial guidance for model selection.

5. Plans for Next Week:

a) Completion of EDA and Dataset Preparation: By next week, we aim to complete the exploratory data analysis and prepare the dataset for model training.

b) Implementation of Machine Learning Algorithm: We plan to implement a Random Forest Regressor model enhanced with XGBoost, following the methodologies outlined in the literature. This will be the primary focus of the coming week.

c) Learning and Implementing Random Forest Regressor with XGBoost: Understanding and coding the Random Forest Regressor with XGBoost might take an additional week beyond the next, but this will position us to effectively build and optimize the model.

6. Materials and Resources:

a) Articles Shared by Faculty:

- A_Hybrid_Approach_for_Interpretable_Game_Performance_Prediction_in_Basketball
- IEEEAccess_JournalPaper1
- NCAA_JournalPaper3
- Senbel_SNPaper
- SpringerNature_JournalPaper2
- SpringerNature_JournalPaper4

b) Additional Online Resources: Various online resources have been referred to for a deeper understanding of the algorithms and techniques discussed in the literature.