# ML Project Weekly Report

**Course name: Machine learning (CSE 523)**

**Week: Week-3 report**

**Group Name: Thunder**

**Instructor's name: Prof. Mehul S Raval**

**Project 9: Data-driven imputation scheme for human-subject based dataset**

## Project progress:

- Conduct of feature sensitivity analysis to identify the most significant features for imputation. This analysis helps prioritize features based on their impact on the imputation process and the overall quality of the dataset.
- Systematic conversion of the data from the provided "Vertical Jump Season 2.xlsx" file, which contains multiple sheets, into a format similar to "Vertical Jump Season 3.csv". This involves organizing the data into a structured format, ensuring consistency in column names, and handling inconsistencies or discrepancies between the two datasets.
- Merging of the formatted data from "Vertical Jump Season 2.csv" with the "Polar.csv" file for Season 2 using the "date" column as the merge key. This step aims to create a single CSV file containing all modalities of data for each athlete, organized date-wise. Care was taken to ensure accuracy in matching athlete names and handling potential data conflicts during the merging process.
- Repeating the merging process for Season 3 data by integrating the formatted data from "Vertical Jump Season 3.csv" with the "Polar.csv" file using the "date" column as the merge key.
- Assessing imputation strategies involving mean, median, and mode approaches for filling missing values, and evaluating their performance in terms of accuracy, computational efficiency, and preservation of the original data's statistical properties.

## Conclusion

- After feature sensitivity analysis is finished, more data processing efforts are guided by the understanding of which features are most important for imputation.

- Compatibility and consistency are ensured through successful conversion and formatting of the datasets, providing the groundwork for efficient analysis.

- Integrating Polar data with Season 2 and Season 3 data allows for thorough analysis over a variety of modalities and time periods.

- Evaluating imputation strategies provides important information about how to handle missing data in a way that balances efficiency, accuracy, and statistical property preservation.