# Project Report: Sensor Data for Behavioural Activity Recognition

# Introduction:

The aim of this project was to utilize machine learning techniques to recognize behavioural activities using sensor data. The project progressed through several stages, including the initial implementation of a random forest model using the iris dataset. Subsequently, the provided data was imported, and seven handcrafted features were extracted, including minimum, maximum, average, standard deviation, and mean values from a single sensor, namely the accelerometer. The initial attempt resulted in a score of 36%.

Day 1: **Random Forest with Iris Dataset**

On the first day, we programmed a random forest model using the well-known iris dataset. This initial step helped us gain familiarity with the implementation and understand the basic concepts of the random forest algorithm.

Day 2: **Handcrafted Features with Single Sensor**

The next day, we imported the provided data and extracted seven handcrafted features from a single sensor, the accelerometer. These features included minimum, maximum, average, standard deviation, and mean values. And got 36% score.

Day 3**: Implementing remaining Handcrafted Features (1)**

Next day, we started implementing the remaining 11 handcrafted features which were not available in numpy, so they have to be implemented. After implementation we tested the model with one sensor data.However, upon testing, we encountered more then 3000 NaN (not a number) values

Day 4: **Using All Sensor data**

After reaching a hurdle of nan values, we decided to spilt paths, half the team member worked on resolving the problem while the other imported the remaining sensor data.

The Team successfully updated the size of feature matrixes to use all sensor data while the other half team was resolving the problem.

Day 5: **Implementing remaining** **Handcrafted Features (2)**

To address Nan issue, we first used Hist Gradient Boosting Classifier, which can handle nan values and used all sensors and got 64% score while using all sensor data. Then we implemented additional checks in our function definition of handcrafted features to prevent nan values to occur and also added imputer with Random Forest so nan values can be resolved. However, the additional checks worked correctly and nan value count was 0. Despite the improvements, the score only increased to 67.7%

Day 6: **Evaluation of each Sensor Score:**

To further refine our results, we used all sensor data one by one to calculate score for each sensor and found that gravity and linear acceleration sensor were not providing good score.

Day 7: **Evaluation of Sensor Combinations**

To further refine our results, we investigated different combinations of sensor feature matrices. We discovered that by using only seven sensors (all except gravity and linear acceleration, we achieved a score of 71%. This finding indicates that a subset of sensors provides optimal information for behavioural activity recognition.

Day 8: **Comparison with Svm**

We used selected sensor data to compare random forest and svm

## Conclusion:

In conclusion, our machine learning project focused on utilizing sensor data for behavioural activity recognition. We progressed through various stages, starting with a random forest implementation using the iris dataset. Subsequently, we extracted handcrafted features from a single sensor, resolving NaN value issues along the way. By incorporating all sensors' feature matrices, we achieved a score of 68%. Further experimentation revealed that using a specific combination of seven sensors improved the score to 71%. These findings showcase the potential of machine learning techniques in recognizing behavioral activities based on sensor data.