### **Next Steps - 10/14/2024**

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### **Universal tracking model for activity - steps, distance, heart rate, predict what type of activity(low, medium high), how accurate model can predict the activity**

### **Datasets with Overlapping Variables:**

1. [**Sleep Health and Lifestyle Dataset**](https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset/data):
   * Heart Rate
   * Sleep Duration
   * Sleep Quality
   * Stress Level
   * Physical Activity Level
   * BMI
   * Blood Pressure
2. [**Apple Watch and Fitbit Data**](https://www.kaggle.com/datasets/aleespinosa/apple-watch-and-fitbit-data/data?select=aw_fb_data.csv):
   * Heart Rate (minute-by-minute)
   * Steps
   * Calories
   * Distance
   * Activity Class (various METS levels)
3. [**Fitbit Data**](https://www.kaggle.com/datasets/arashnic/fitbit):
   * Heart Rate
   * Total Steps
   * Distance
   * Sleep Duration
   * Activity Levels (Very Active, Moderately Active, Light Activity, Sedentary)
4. [**Fitbit Sleep Score Data**](https://www.kaggle.com/datasets/mbalos/fitbit-sleep-score-data):
   * Sleep Score
   * Deep Sleep Minutes
   * Resting Heart Rate
   * Restlessness
   * Overall Sleep Quality
5. [**Sleep and Health Metrics**](https://www.kaggle.com/datasets/uom190346a/sleep-and-health-metrics):
   * Sleep Quality
   * Sleep Duration
   * Heart Rate Variability
   * Stress Level
   * Caffeine Intake

### **Common Variables Across Datasets:**

1. **Heart Rate**: Present in all five datasets, capturing either resting heart rate or detailed heart rate variability.
2. **Sleep Duration/Quality**: All datasets have sleep metrics, including duration, quality scores, or sleep stages.
3. **Stress Level**: Directly available in some datasets, and can be inferred from heart rate variability in others.
4. **Physical Activity**: Steps, active minutes, or activity class data are present in most datasets.

### 

### **2. Heart Health Prediction and Monitoring**

* **Objective**: Create an ML model to predict potential heart health issues using heart rate, blood pressure, and stress data.
* **Datasets to Use**:
  + [HRV Data](https://www.kaggle.com/datasets/qiriro/swell-heart-rate-variability-hrv)
  + [Heart Beat ECG Data](https://paperswithcode.com/dataset/ecg5000)
  + [Fitbit Data for Heart Rate and Steps](https://www.kaggle.com/datasets/arashnic/fitbit)
* **Key Variables**: Heart rate variability, blood pressure, stress level, physical activity, sleep quality, and BMI category.
* **Potential Insights**: The model could identify correlations between stress levels, heart rate, and activity patterns, and provide insights into lifestyle changes to improve heart health. It could also alert users to potentially dangerous heart rate variations based on their data.

### **Implementation Plan**

* **Data Collection & Labelling**: Aggregate and preprocess the relevant datasets, ensuring they are properly labeled for supervised learning tasks.
* **Model Development**: Choose algorithms like random forests, neural networks, or gradient boosting for prediction tasks. For time series data, consider LSTM or GRU models.
* **Evaluation Metrics**: Use appropriate metrics like MAE, RMSE for regression tasks (e.g., predicting sleep duration), and accuracy, F1-score for classification tasks (e.g., predicting stress level).
* **Recommendation Engine**: Based on the model outputs, develop a rule-based or ML-based system to provide actionable insights and recommendations.

**### 2. \*\*Activity Recognition\*\***

- \*\*Model:\*\* A classification model (e.g., Random Forest, SVM, Deep Learning) that uses heart rate, step count, and movement data to predict user activity (e.g., walking, running, sleeping).

- \*\*Data Sources:\*\*

- [MHealth Dataset](http://archive.ics.uci.edu/ml/datasets/MHEALTH+Dataset): Includes motion and physiological signals.

- [Activity Recognition from Single Chest-Mounted Accelerometer Data](https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+from+Continuous+Ambient+Sensor+Data)

**### 4. \*\*Calorie Burn Estimation\*\***

- \*\*Model:\*\* Regression model (e.g., Linear Regression, Neural Networks) using heart rate, activity level, and demographic data (age, weight, gender) to estimate calorie burn.

- \*\*Data Sources:\*\*

- [Fitbit Data on Kaggle](https://www.kaggle.com/datasets/): Contains user activity, heart rate, and calories burned.

- [Google Fit API](https://developers.google.com/fit): Allows access to various health data metrics.