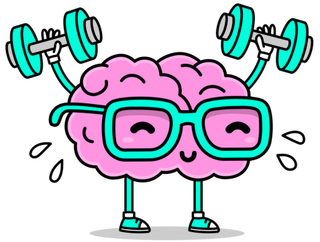
HEALTH
TRACKER DATA

CEM SÜMER

Sabancı University
Junior CSE Student
dreamer



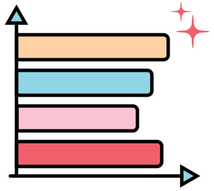
MOTIVATION

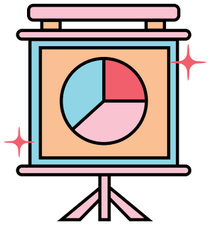
When I first started the project, the term "own data" seemed very restrictive and challenging to me. Then I listed the fields that belong to me and from which I can extract data. Since I have been doing sports for a long time, I thought that I might have different health trucker data than a normal person, and I was excited to take action to process my data. I wanted to do a health trucker data project because I was curious about the subtle differences, developments and changes between when I do sports and when I don't.

DATA SOURCE

In this section, I will try to explain in steps where and how I obtained the data. Frankly, obtaining the data was much more challenging than I thought. To get my Apple(ios) Health Trucker data, I first requested it from Apple's own application. When it was integrated into my phone with xml extension, I was able to obtain all the data it captured simultaneously. I used ready-made Python libraries to convert this data, which was in xml format, into csv format and I was able to turn it into a single csv file. All of this was challenging and instructive.

DATA ANALYSIS





CONTENTS

Utilized Dataset

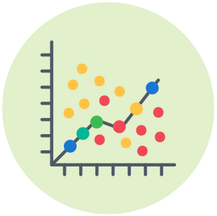
Feature Engineering

Machine Learning Models

Lineer Regression

XGBoost

kNN



Utilized Dataset

- **Data Loading and Overview**
- **Heart Rate Statistics**
- **Column-wise Statistics (Excluding”Heart Rate and “Date”)**
- **Categorizing Heart Rates**
- **Label Encoding Heart Rate Categories**

Feature Engineering

- **Correlation Analysis: Between Step Count and Active Energy Burned**
- **Hypothesis Testing Results**
- **Step Count Histogram**
- **Active Energy Burned Histogram**
- **Creation of Activity Intensity Feature**

Lineer Regression

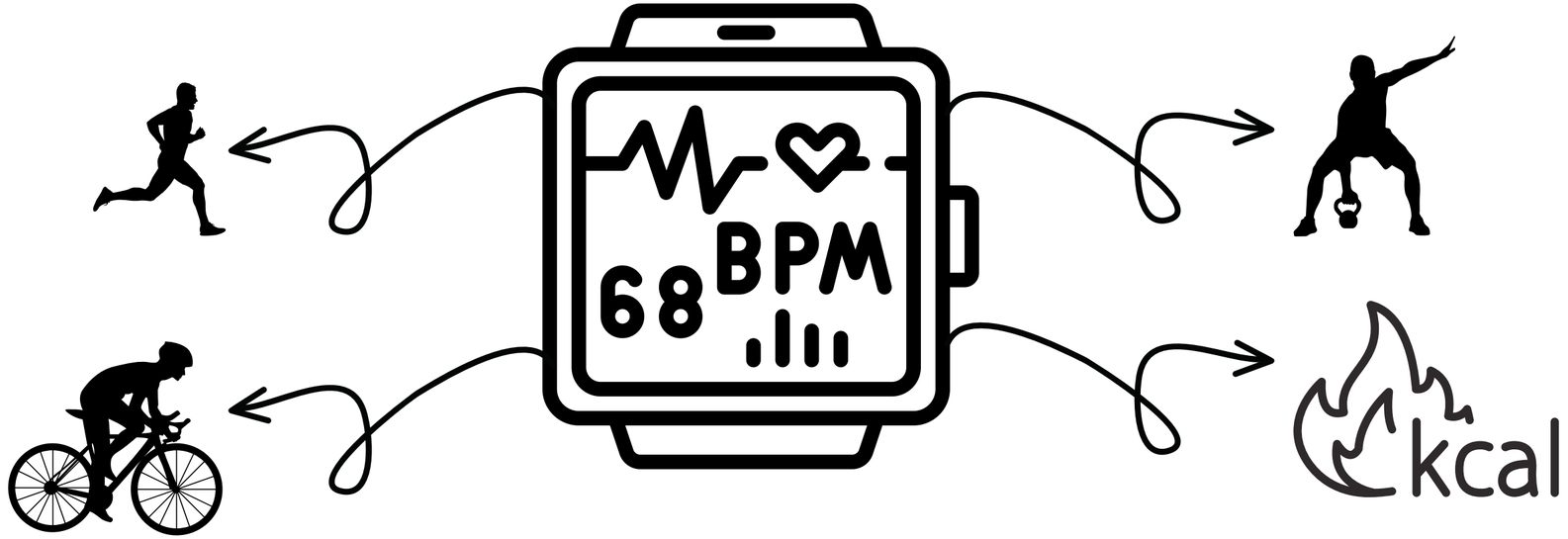
XGBoost

kNN

- **Linear Regression Model Training and Evaluation**
- **Linear Regression Model Predictions on Test Set**
- **Prediction Using Linear Regression on New Data**
- **XGBoost Model Tuning and Evaluation**
- **XGBoost Model Predictions on New Data**
- **k-Nearest Neighbors (kNN) Model Tuning and Evaluation**
- **kNN Model Predictions on New Data**
- **Comparison of Predictions Between Models**

FINDINGS

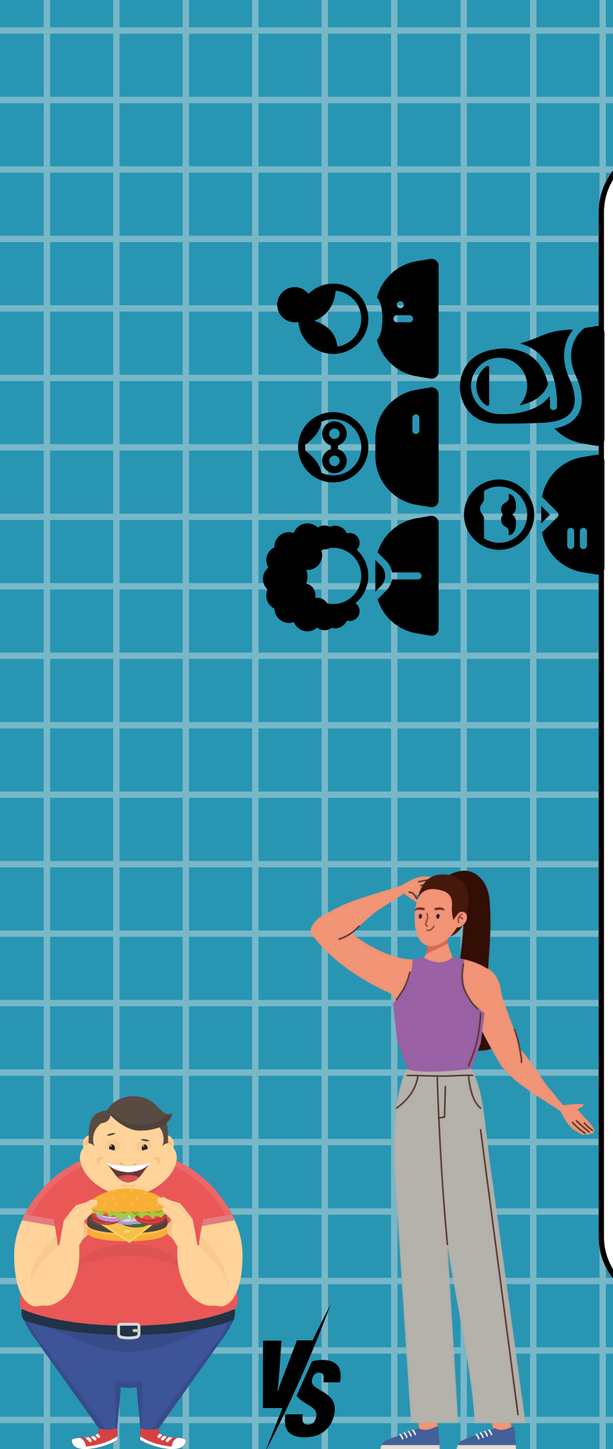
I learned that I can estimate my heart rate by looking at my own daily activities (running, walking, doing sports, cycling, sleeping, resting) and how to classify it. I have the idea of what the correlation between daily activities and heart rate is.

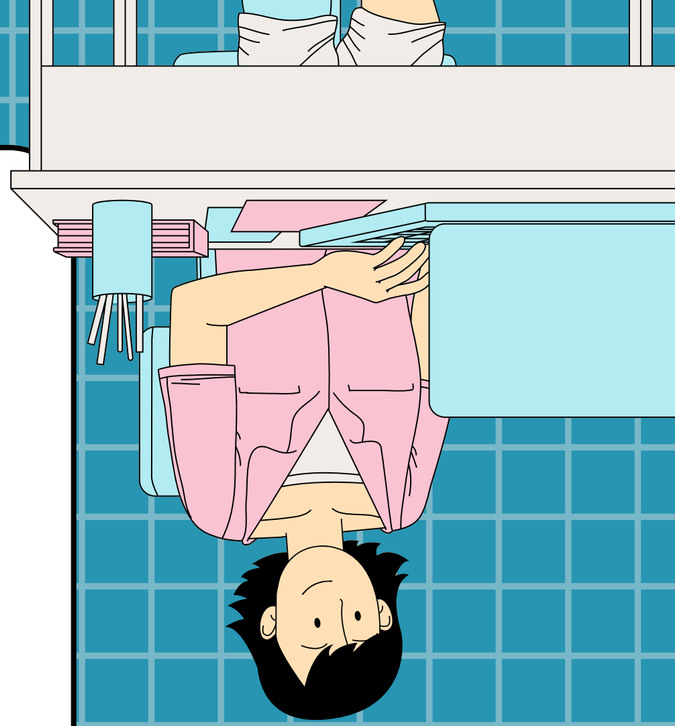


LIMITIATIONS AND FUTURE WORK

By taking the data of other friends (for example, friends with different physical characteristics) and comparing them with each other, I can create much more complex models and analyzes with a broader perspective.

It was a simple model because I used only my own data, that is, a single person. It is a bit difficult to put forward complex models with limited data because the data diversity is low.





"Data dance complete!
Here's to decoding the rhythm of insights and
waltzing into a future of boundless possibilities.
Thanks for grooving with the data vibes!"