Sleep, and Health

Team 1
Chen Yang (002837912)
Zihao Lu (002642258)
Pei-Han Hsu (002244953)

CONTENT

- Use cases
- Methodology
- Data sources
- Milestones/sprints
- What will you program in Scala
- Acceptance criteria
- Goals of the project

USE CASE

- Program runs on the play framework interface
- User inputs the data the system display the prediction on another tab
- Program received csv data from backend

METHODOLOGY

- Investigate the relation and trend between sleep conditions and health, provide predictions for people how is their health
- Provide an interactive interface for the users to explore about sleep and health
- **Sleep Conditions** include(these would be put inside vector)
  a. Sleep duration
  b. Sleep quality (mark as scores)
  c. Stress Level
- **Health** (y-outputs)
  a. Blood pressure
  b. BMI number

DATA SOURCES

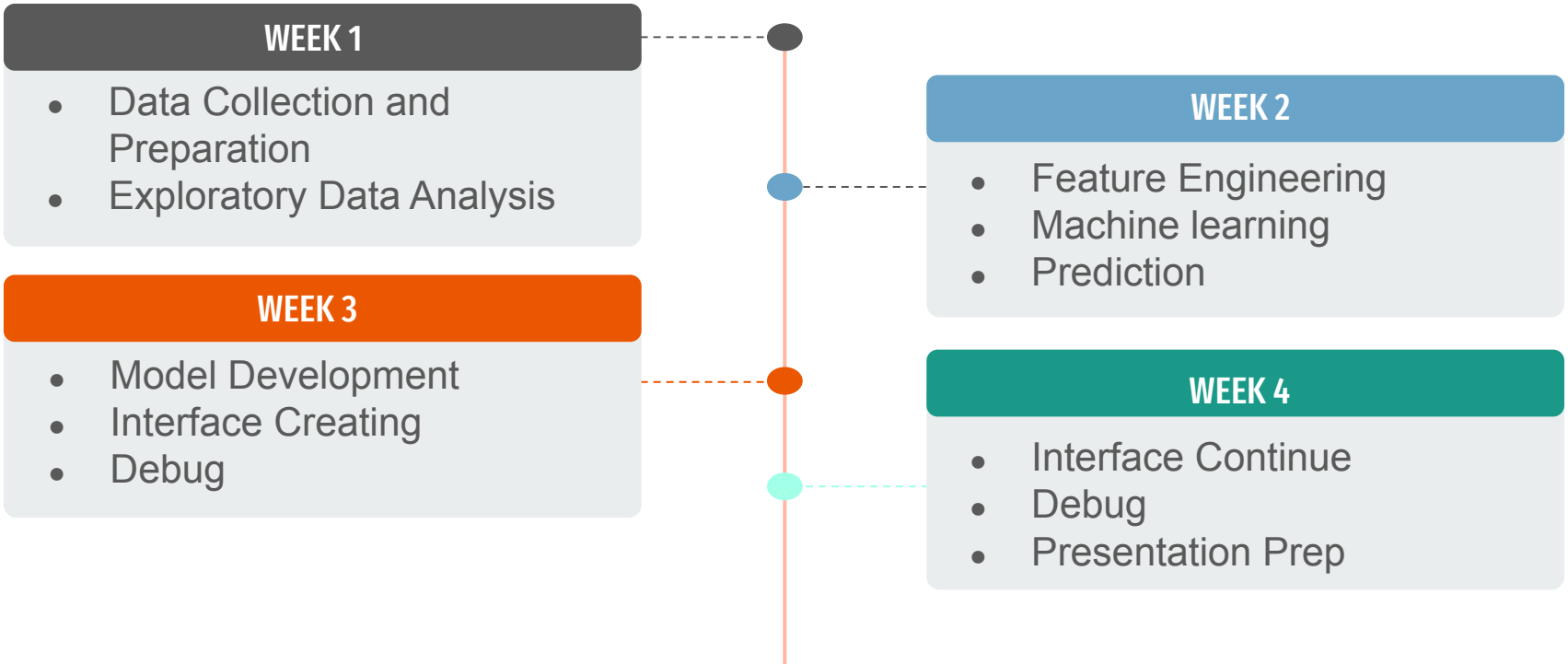
- Sleep Health and Lifestyle Dataset

  373 rows, 13 columns, {daily routines, gender, age, profession, sleep duration, sleep quality, physical activity, stress levels, BMI classification, blood pressure, heart rate, daily step count, and presence of sleep disorders}.

- Heart Attack Analysis & Prediction Dataset

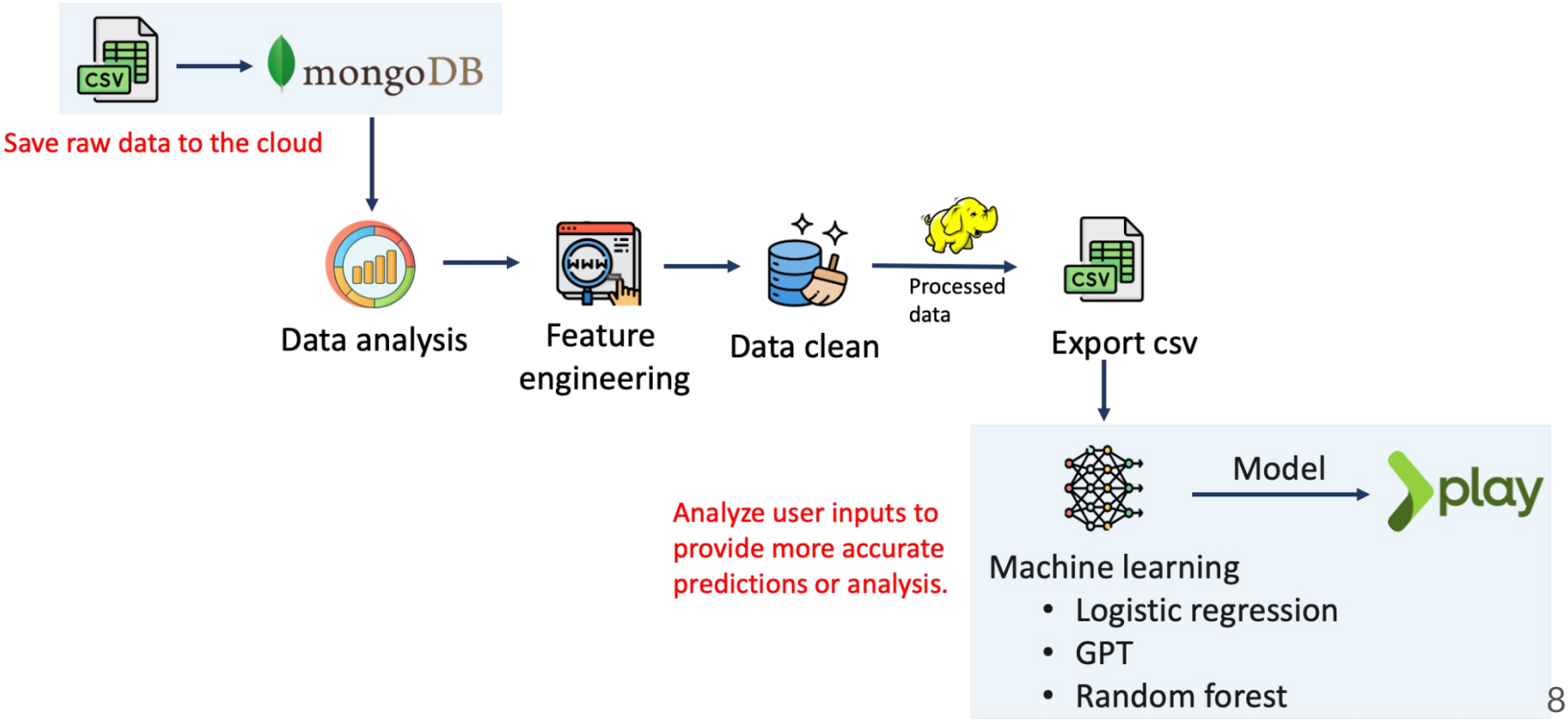
  303 rows, 15 columns, {age, sex, exercise-induced angina, number of major vessels, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, heart rate achieved}.

MILESTONES/SPRINTS



WHAT WILL YOU PROGRAM IN SCALA

- Data Analysis for each dataset (csye 7200final/app/Model/DataAnalysis.scala)
- Featuring Engineering (csye 7200final/app/Model/MongoDBDF.scala)
- Machine learning methods (Logistic Regression, Generative Pre-trained Transformer (GPT)) (csye 7200final/app/Model/GBT.scala and Logitsic.scala)
- Play framework with scala to build controller combined with html (csye 7200final/app/HomeController.scala)
- MongoDB (atlas cloud) with scala to get the data remotely(csye 7200final/app/Model/MongoDBDF.scala)
- Code Location: https://github.com/ZihaoLu1106/CSYE7200-teamHSYL-Final


CSV
mongoDB
Save raw data to the cloud
Data analysis
Feature engineering
Data clean
Processed data
CSV
Export csv
Analyze user inputs to provide more accurate predictions or analysis.
Model
play
Machine learning
• Logistic regression
• GPT
• Random forest

ACCEPTANCE CRITERIA

- **Before Programming**

  Prediction Accuracy : 70%

- **Actually**
  - gbt: 56%
  - Logistic: 62%

So we decided to use Logistic, base on its **higher accuracy** and **higher speed.**

GOALS OF THE PROJECT

- Goal 1: To investigate the relation between sleep (with its condition) and BMI and blood pressure. **(checked)**
- Goal 2: Investigate our self-sleep hours and behaviors as inputs, to predict our BMI and blood pressure. **(checked)**
- Goal 3: Provide bunch of functions in the interface for the users to investigate their own sleep and health conditions. **(checked)**

DEMO