CP630 Project Proposal
Detect a Heart Condition with Machine Learning
Josh Barber

2021-11-30

1. Introduction

Cardiovascular disease ranks first in the causes of death globally, taking an estimated 17.9 million lives per year. That accounts for 31% of all deaths worldwide. People with cardiovascular disease need early detection and introducing a machine learning algorithm to help determine if said person has it can be beneficial. The algorithm can determine cardiovascular disease on 11 questions.

The main goal of this project is to determine if said user may have a cardiovascular disease by using machine learning techniques by the Weka API on a dataset from Kaggle.

2. Problem solving and algorithms

Application data

- **data representation and dataset design**

    The data representation can be viewed as a classification model, where outcome belongs to two labels, "Yes" or "No" of having a cardiovascular disease. The dataset design includes 11 other columns of preliminary questions including age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting ECG, max heart rate, exercise angina, oldpeak, and the slope of the peak exercise. All data is complete and has 1190 observations from five different datasets combined making it the world's largest heart disease dataset available. There are 272 duplications, so this dataset only contains 918 observations of the 1190 total observations.

- **data collection and pre-processing**

    The data will be collected via website form and will use a Weka classification algorithm to help predict a result. The classification model will initially be stored in the database and will be fetched before returning a result to the user after their form submission. The results from the user will be collected and stored in the database to continue training the classification algorithm. The data will be supervised, so all data will be complete on form submission.

Models to represent information, knowledge and patterns.

There will be one model that will constantly update the database on user input.

**Algorithms to solve the problems and compute the models.**

The algorithms used for the classification model can be done by linear regression or decision trees. The admin can decide to change the algorithm in the admin portal.

3. Proposed System Design

System design of your solutions as an enterprise computing application.

User can access the website from domain lookup (in this case localhost). Will prompt a questionnaire of 11 questions to determine if they may have cardiovascular disease. The user can then submit the form to get results. Admin user can login into the administrator portal to change Weka models from regression to decision tress if admin so chooses to.

System architecture (better use a diagram to illustrate the components and their relations).
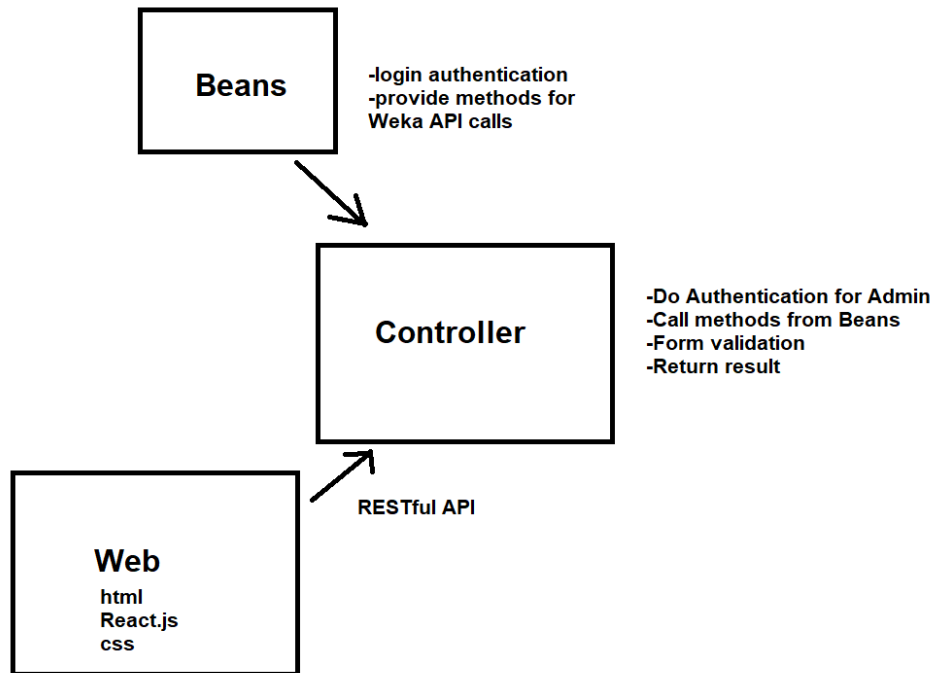
Platform and tools to be used in the project.

- Eclipse
- MySql
- Java EE
- Weka API
- React.js

4. Project plan and schedule

List of tasks/milestones/check points of your project with time schedule. For team project, it needs to provide the roles and tasks of each member.

Task	Week 11	Week 12	Week 13	Week 14
Build Session Beans	timeline			
Build Web		timeline		
Build Controller			timeline	
Presentation				timeline

5. References

A list of references you read for your project. papers, articles, available datasets repositories.

1. Datasource: https://www.kaggle.com/fedesoriano/heart-failure-prediction
2. Weka 3: Machine Learning Software in Java. https://www.cs.waikato.ac.nz/ml/weka/