Lean Analytics for a Leaner Person

Hello!

We are Group #8

Manuel Maldonado
Ming Tai Ha
Nick Taylor
Tongpen Zhang
William Karavites

1

Data Collection & Model

How did we collect our data and build our model

Why CDC over Twitter?

- Systematic methodology for collecting statistics
- More structured information present in CDC data
- Data present allows for more ambitious analytics
- Format enabled use of ML to extract knowledge.

"National Health and Nutrition Examination Survey" - Annual survey conducted by the CDC to assess the health and nutritional status of people in the United States (2007-2012 data used).

Model

- A Regression Tree classifies data based on a chosen set of attributes
- This model is susceptible to overfitting and requires pruning
- Tree Structure: This tree uses N attributes, each attribute i with Ki choices:

Root
Attribute 1, choice 1
...(more attributes)
Attribute N, choice 1
...
Attribute N, choice KN
... (more choices)
Attribute 1, choice K1
...

Current Algorithm:

```
if no attributes left to classify:
    pass
if represents less than 10 data points:
    pass        //    Static Pre-Pruning
else
    find new attribute that minimizes the average range
    of the statistic of subsetted data
    if no new attribute has average range smaller than the
      range of the statistic current data
else
        Create children nodes for each choice
        Recurse for each node
```

2

Datastore

The link between the Health Model, Processor and Database

Datastore

Divided into two sub-systems: Driver and Data Ops

Driver acts as a data abstraction layer containing all PHP-PDO functions

Data Ops accepts commands from the Processor and contains all SQL queries

The Model and Datastore communicate via JSON files

This file is exported by the Model and imported into the database through the Datastore.

The reverse process occurs when passing new data to the Model

The database management system used is MySQL.

The team has much more experience MySQL in comparison to other DBMS

Future: Algorithm that converts the JSON file into records saved in the database

3

Processor

The link between the Health Model, Processor and Database

The Processor

What is it?

In its broadest definition, it is a PHP server-side application, that receives, handles, processes requests sent by the user interfaces.

What does it do?

It implements a RESTful API that the website and the mobile application can use to send/receive data. It connects users and their dats.

Future:

Still needs implementation. We focused on the Model, Website and Mobile App.

4

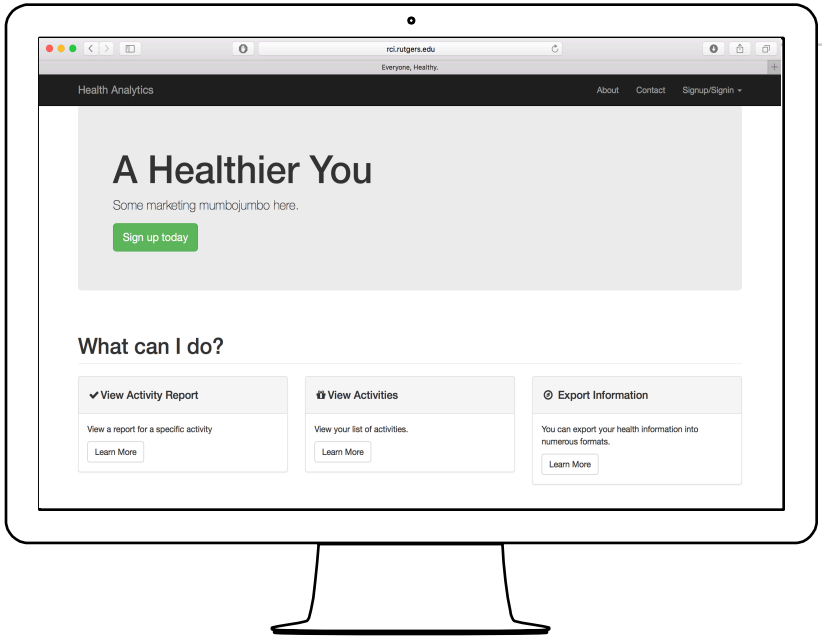
Website and Mobile Application

How our users will interact with the system

Website project

Created using Bootstrap, AngularJS & the Google Maps API.

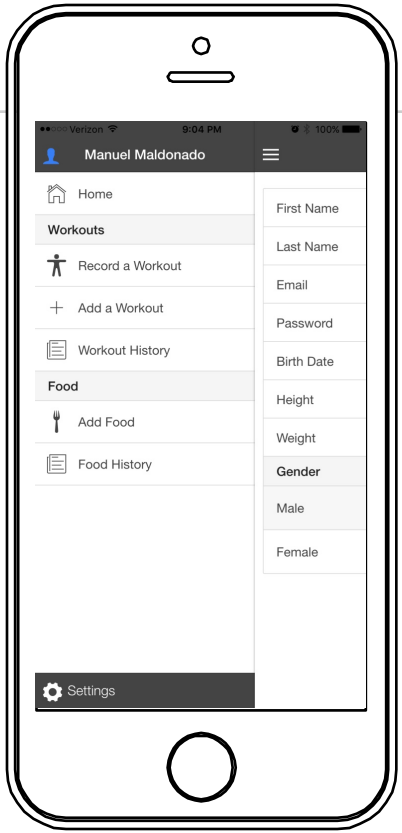
These tools allow for efficient MVC design, specifically AngularJS.



Mobile project

Created using IonicFramework, AngularJS and Apache Cordova.

These tools also allow for great MVC design and cross-platform support.



Thanks!

Any questions ?

You can find our project information at:

- Website: http://willkara.com/projects/HealthAnalytics/about.html
- Blog : http://blog.willkara.com/tag/software-engineering-class-project/
- Full source code will be pubilicaly available shortly.