Georgia Tech Students’ work building PLP UI

ABHISHEK KHOWALA
CS 6440 – Introduction to Health Informatics course give students opportunity to work on real world high impact healthcare related projects in team supported by External Mentor and TA Mentor

~15 weeks long project

Typical team size of 4-6 students

The delivery of project is broken into 6 milestones

Infrastructure for continuous build and deployment provided

FHIR servers provided
Deliverables

- D0 – Team formation & Team Confirmation (WW03)
- D1 – Project Topic: Team Topic, Research, Use Cases, Project Plan (WW06)
- D2 – Technical Presentation: Technical Architecture (WW09)
- D3 – Team Progress: Effectively show elements of application (WW12)
- D4 – Full demonstration of the application (WW15)
- D5 – Final Project Submission (WW15)

- Students provide mentors with weekly update irrespective of any major delivery or not
- Each deliverable includes presentation as well as a narrated video link, every student in the team need to participate
DEPLOYMENT INFRASTRUCTURE FOR CONTINUOUS BUILD, CONTAINERIZATION AND DEPLOYMENT INTO A SERVER FARM
HDAP – Health Data Analytics Platform

- Software platform designed to support students in conducting health oriented projects at Georgia Tech
- HDAP role in projects
  - Access to FHIR servers
  - Application Deployment
- HDAP 3 core functions
  - Provide synthetic and de-identified sources of healthcare data
  - Provide tools to analyze these data
  - Provide a healthcare application development and hosting environment and app galleries
HDAP – Continued:

- FHIR SERVER ACCESS PROVIDED FOR
  - **MIMIC (STU3)**
    - OMOP data
    - **Endpoint:** https://apps.hdap.gatech.edu/gt-fhir/tester/
  - **SyntheticMass (STU3)**
    - Endpoint: https://apps.hdap.gatech.edu/syntheticmass/
  - **HAPI (R4)**
    - Endpoint: https://apps.hdap.gatech.edu/hapiR4/
Development/Deployment Pipeline

- **Student** pushes to **GitHub**
- **GitHub** pulls and notifies **Jenkins**
- **Jenkins** pushes to **Docker Registry**
- **Docker Registry** pulls and notifies **Rancher**
- **Rancher** pushes to **Load balancer**
- **Load balancer** forwards to **Apache**
- **Apache** processes the **HTTPS request**
- **Reverse proxies** route to **Jenkins**

---

**Development**

- **Student**
- **GitHub**
- **Jenkins**
- **Docker Registry**
- **Rancher**
- **Load balancer**
- **Apache**

**Deployment**

- **Student**
- **GitHub**
- **Jenkins**
- **Docker Registry**
- **Rancher**
- **Load balancer**
- **Apache**
- **Reverse proxies**
Process

- Student pushes code to GT GitHub
- GT GitHub Notifies Jenkins of push
- Jenkins builds project and pushes images to the private Docker registry
- Jenkins notifies Rancher of the push to the registry
- Rancher pulls the images, builds containers, and deploys them to the Docker Swarm
OHDSI PLP UI Project

FALL 2018 SEMESTER: CONSUMER FACING APPLICATION FOR PATIENT LEVEL PREDICTION
TEAM COMPOSITION

TEAM MEMBERS
- MICHAEL GARNER
- JONATHAN FUNG
- ALYSSA DE LEON
- JACOB GILBERT
- CHUREN (CHU) SHAO

TA MENTOR
- ABHISHEK KHOWALA

EXTERNAL MENTOR
- DR. PATRICK RYAN
- DR. DAVID MADIGAN

HEAD TA
- TIA POPE

PROFESSOR
- DR. JON DUKE
PROJECT GOALS

- Analyze 3 example OHDSI PLP Models viz. AfibStroke, DepressionSuicide, NSAIDGIbleed
- Create a single interface that takes any PLP results object (such as 3 examples provided) and serve up an ordered set of questions to allow a target population user to estimate a personalized prediction of their risk of the outcome
- Use coefficient and covariate summary together with user answers to yes/no questions to produce the predicted probability and estimated confidence interval (CI) around those predictions.
# Project Planning: Gantt Chart

## Deliverable - 1: Team Topic Presentation

<table>
<thead>
<tr>
<th>Task Title</th>
<th>Task Owner</th>
<th>Start Date</th>
<th>Due Date</th>
<th>% of Task Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>Michael D'Her</td>
<td>10/11/16</td>
<td>10/15/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 2</td>
<td>Research</td>
<td>10/15/16</td>
<td>10/17/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 3</td>
<td>freshers Class</td>
<td>10/15/16</td>
<td>10/17/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 4</td>
<td>Introduction &amp; Consultancy</td>
<td>10/15/16</td>
<td>10/17/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 5</td>
<td>Maratoni</td>
<td>10/15/16</td>
<td>10/17/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 6</td>
<td>Saliqon</td>
<td>10/15/16</td>
<td>10/17/16</td>
<td>100%</td>
</tr>
</tbody>
</table>

## Deliverable - 2: Technical Presentation

<table>
<thead>
<tr>
<th>Task Title</th>
<th>Task Owner</th>
<th>Start Date</th>
<th>Due Date</th>
<th>% of Task Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>All</td>
<td>10/15/16</td>
<td>10/17/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 2</td>
<td>Architectural Diagram</td>
<td>Jonathan and Jacob</td>
<td>10/20/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 3</td>
<td>Presentation</td>
<td>10/18/16</td>
<td>10/20/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 4</td>
<td>Maratoni</td>
<td>10/18/16</td>
<td>10/20/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 5</td>
<td>Presentation Update</td>
<td>10/20/16</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 6</td>
<td>Meeting with Mandus</td>
<td>10/20/16</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 7</td>
<td>Suborsi</td>
<td>10/20/16</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
</tbody>
</table>

## Deliverable - 3: Progress Report

<table>
<thead>
<tr>
<th>Task Title</th>
<th>Task Owner</th>
<th>Start Date</th>
<th>Due Date</th>
<th>% of Task Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>Research (Questions and Model)</td>
<td>All</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 2</td>
<td>Setup Development Environment</td>
<td>Chenne</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 3</td>
<td>Front End Development</td>
<td>Jonathan and Alyssa</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 4</td>
<td>Back End Development</td>
<td>Michael and Chunes</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 5</td>
<td>Testing and Q.A.</td>
<td>Jonathan and James</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 6</td>
<td>Progress Presentation</td>
<td>Alyssa and Chunes</td>
<td>10/22/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 7</td>
<td>Presentation Update</td>
<td>10/22/16</td>
<td>10/24/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 8</td>
<td>Maratoni</td>
<td>10/22/16</td>
<td>10/24/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 9</td>
<td>Meeting with Mandus</td>
<td>10/22/16</td>
<td>10/24/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 10</td>
<td>Jomino</td>
<td>10/22/16</td>
<td>10/24/16</td>
<td>100%</td>
</tr>
</tbody>
</table>

## Deliverable - 4 and 5: Front End Development

<table>
<thead>
<tr>
<th>Task Title</th>
<th>Task Owner</th>
<th>Start Date</th>
<th>Due Date</th>
<th>% of Task Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>Jonathan and Alyssa</td>
<td>11/13/16</td>
<td>11/23/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 2</td>
<td>Front End Development</td>
<td>Michael and Chunes</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
<tr>
<td>Task 3</td>
<td>Testing and Q.A.</td>
<td>Jonathan and Jacob</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
<tr>
<td>Task 4</td>
<td>Manual</td>
<td>Jacob</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
<tr>
<td>Task 5</td>
<td>Special Instructions</td>
<td>Jonathan</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
<tr>
<td>Task 6</td>
<td>Presentation</td>
<td>Alyssa and Chunes</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
<tr>
<td>Task 7</td>
<td>Maratoni</td>
<td>Jacob</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
<tr>
<td>Task 8</td>
<td>Presentation Update</td>
<td>11/13/16</td>
<td>11/23/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 9</td>
<td>Meeting with Mandus</td>
<td>11/13/16</td>
<td>11/23/16</td>
<td>100%</td>
</tr>
<tr>
<td>Task 10</td>
<td>Delivery</td>
<td>Chunes</td>
<td>11/13/16</td>
<td>11/23/16</td>
</tr>
</tbody>
</table>
Workflow

- Choose Prediction Model of Interest
- Create Questions from given medical codes used in model
- Order questions by impact to prediction (reordering if necessary)
- Present questions to user
- Update predicted outcome and confidence level as questions are answered
Architecture

1. Select Target Cohort (which bucket the user belongs in)
2. Select Outcome Cohort for a related disease (which bucket the user wants to know more about)
3. Select Time Interval (the chance that outcome will occur in next <time interval>)
7. Prompt user for question showing current predictions after each question

4. Generate predictive model based on previous selections using OHDSI Common Data Model Data
5. Generate questions based off of covariates used to train model
6. Sort Questions by covariate effectiveness in model. Re-sort after each question
Design

- Read R models (.rds)
- Process Covariate data
- Sort Questions
- Perform Monte Carlo Estimations
- Extendible: Ability to add new PLP models through simple configuration steps
Sorting Covariate Questions

- The top equation determines the weight of evidence.
- The bottom equation is for determining the expected weight of evidence on a specific health state.
- Questions are sorted by descending expected weight (EW) value.

\[ W(H : E) = \log \left( \frac{Pr(E|H)}{Pr(E|\overline{H})} \right) \]

\[ EW(H : E) = \sum_{k=1}^{m} W(H : t_k) Pr(t_k|H) \]
Monte Carlo Estimation

Steps to complete Monte Carlo Estimation:

- Simulate the missing covariate data by covariates marginal probability
- Run logistic regression on simulated covariate outcome from step 1
- Stores outcome and repeats steps 1 and 2 a total of ~2000 times
- Returns mean and standard deviation of stored outcomes. These are used to calculate the 95% interval.
Development Phase

- Developed in Django
- 3 Pre-trained R models from OHDSI
- Rpy2 library
- Backend and Frontend Development

Select an Outcome

- AfibStrokeModelTransport (prob: 0.5932617187499998)
- DepressionSuicideModelTransport (prob: 0.3955078124999999)
- NSAIDGIBleedModelTransport (prob: 0.3)

DepressionSuicideModelTransport

Question: 1/280

Have you ever been diagnosed with and/or suffered from “Major depressive disorder”?

- Yes
- No
- Don’t Know

Prediction with 3.04% chance - 10.82% (within 95% confidence interval)
Polish Phase

- Gathering Feedback
- Round 1: too many pages
- Round 2: too many questions
- Added css and js
Testing Phase

- Bug fixing
- Storing sessions

Ischemic stroke for patients diagnosed with atrial fibrillation

Amongst patients newly diagnosed with atrial fibrillation, predict the probability the patient will have an ischemic stroke within the next 3 years.

Please answer the following questions. Take a look at your prediction and standard deviation. The more questions you answer, the lower your standard deviation will become. Hence, more accurate the prediction will be.

11. Have you ever been diagnosed with and/or suffered from peripheral vascular disease?
   - Yes
   - No
   - Unsure

12. Have you ever been diagnosed with and/or suffered from fatigue?
   - Yes
   - No
   - Unsure

13. Have you ever been diagnosed with and/or suffered from general problem and/or complaint?
   - Yes
   - No
   - Unsure
DEMO

HTTPS://CS6440-F18-PRJ14.APPS.HDAP.GATECH.EDU/
Project Public URL: https://cs6440-f18-prj14.apps.hdap.gatech.edu/

PLP objects description: https://github.com/OHDSI/PatientLevelPrediction

3 example PLP models: https://drive.google.com/drive/folders/1wcRsUIAIXXsUMLRLW_vP-slXz7bzL8V?usp=sharing

Gantt Chart: https://docs.google.com/spreadsheets/d/1FNK3MJTs4e05rsHUXeUyAzfJcM83Is52t-Lmf0Ywhqo/edit#gid=0
Links/References: Continued

- PLP model JAMIA paper: https://academic.oup.com/jamia/article/25/8/969/4989437
- Project Git repository: https://github.gatech.edu/gt-cs6440-hit-fall2018/Consumer-facing-application-for-patient-level-prediction
Thank You!

Questions?