Data-Driven Precision Healthcare Analytics Team Charter & Project Plan

I. Team Charter

(a) Team Roster

Name and Contact	**Zezhi Zhang** **510-944-9774** **zezhi_zhang@berkeley.edu** **https://www.linkedin.com/in/zezhi-zhang-b89a7418a/**	**Aparna Mohan** **408-518-9911** **aparna_mohan@berkeley.edu** **https://www.linkedin.com/in/aparna-mohan1/**	**YiYi Chen** **510-984-8501** **yiyichen@berkeley.edu** **https://www.linkedin.com/in/yiyi-chen-044aa1105/**
Current Tech Expertise	Statistical knowledge and tools and computer skills	Project management and data analysis	Industrial Engineering and Computer Science background
Desired Technical	Develop knowledge of python libraries like numpy,matplotlib and pandas and supervised machine learning algorithms like Logistic Regression.	Strengthen Python programming skills, data analysis, visualization and machine learning algorithms like CART.	Apply Pandas library and Matplotlib to visualize data frames, analyze datasets by Random Forest and help the deliverables on time by using project management skills.
Technical Focus	Zezhi will leverage his mathematical background to find out the method to build the model.	Aparna will assist in model selection and training the dataset.	Yiyi will do the coding part and make sure the capstone project is delivered on time by applying project management skills.
PM Responsibility	Point-of-Contact/Communication Manager	Team Lead	Project Manager

(b) Advising Roster

Name and Contact	**Anil Aswani** **510-664-9114** **aaswani@berkeley.edu**
Advising Role	Faculty Liaison
Affiliation	Associate Professor, Industrial Engineering and Operations Research Department

(c) Communication Plan

Team INFORMATIONAL Meetings

- *Location:* Shires Hall; Zezhi will be responsible for room bookings; Aparna will act as a backup.
- *Time:* Mon, 2-3pm
- *Duration:* 1hr
- *Frequency:* weekly
- *Agenda:* Aparna

Team WORK Meetings

- *Location:* Shires Hall; Zezhi will be responsible for room bookings; Aparna will act as a backup.
- *Time:* Wed, 10 am-1 pm
- *Duration:* 3hrs
- *Frequency:* weekly; later in the term, may be adjusted to twice per week
- *Agenda: Yiyi*

ADVISING Meetings

- *Location:* 4119 Etcheverry Hall
- *Time:* Mon, 1-1:30pm
- *Duration:* 0.5hr
- *Frequency:* once a week.
- *Agenda: Zezhi*

Remote Meetings

- *Internal:* While in-person meetings are preferred, we can meet virtually if needed (using zoom/google calendar); in case someone can't make the meeting, they will contact Aparna who will arrange a zoom invite. Zezhi will act as Aparna's backup.
- *External:* We will have a weekly virtual meeting with our stakeholders (using the zoom/google calendar). Aparna will be arranging these meetings. Zezhi will act as Aparna's backup.

Emergencies:

- If someone can't make our regular in-person meeting, they will contact Aparna at least 2hrs prior to the start of the meeting. Aparna prefers a phone call (please don't use email or text messaging)
- To ensure everyone's health: please call in remotely if you're sick.

(d) Project Management Tool

We'll use EduSourced to assign and track tasks.

II. PROJECT PLAN

Project Vision & Objective	**Interim Deliverables**	**Associated Tasks / Leads**	**Final Deliverable**
Our team will analyze the data from NHANES and develop algorithms to predict the risk for diabetes and depression into the home setting by using machine learning and other data analytics tools. By Spring of 2020, we will create a model that can predict whether a person is at risk of diabetes and depression according to patient's health and lifestyle data.	Select data set, filter the independent variables and visualize the data to understand the correlation. Now to 10/15/19	Aparna will lead in selecting the Real healthcare data sets about physical activity and exercise trajectories, chronic disease states and progression, dietary consumption decisions and behaviors, etc. The team will learn to use pandas library to transfer the data set into python dataframe. They will also visualize the relationship between the disease and some filtered variables from the NHANES.	Prediction model to assist in the early detection of diabetes and depression from medical history, lifestyle and dietary data.
	Select the analysis method Now to 11/8/19	The team will learn several analysis methods and algorithms like random trees, support vector machine and logistic regression and select the algorithm most applicable to the dataset. This will be led by Zezhi.	
	Build the prediction model to 12/1/19	Once the data set and analysis method are decided, our team will build the prediction model from the data and test its predictive accuracy with the data. Aparna will take the lead.	
	Apply recommendations to revise the model By 2/1/20	Discuss the model with professor and apply his advice and recommendation to the model.	
	Data valuation By 2/15/20	Measure the accuracy of the model and improve its performance. The success metric is the accuracy of the model and reflection of the model. Highlight the most influential predictors. Give several possible directions for future research.	
	Finish the report 2/15/20 to 3/15/20	This work will be led by Aparna who will document the results of our analysis	
	Finish all the work 3/15/20 to 4/15/20	Yiyi will lead this section to ensure the completion of documentation and supplement, if necessary.	