Portland State
UNIVERSITY

To: Dr. Mara Tableman

From: Bruce Marron

Subject: Consulting Update 1

Date: January 5, 2017

Thank you so much for structuring our class in a "get out there and get the job done" approach. I find such an approach pragmatic, professional, and truly enjoyable.

As previously mentioned, I will be working with Alejandro Queral, Director Systems Planning and Performance at United Way of the Columbia-Willamette. Relevant information about him and his role at United Way is available at his website, `www.unitedway-pdx.org/about-us/team/our-staff/alejandroq`. I will be sure to cc him on the transmittal of this update so that you have his email directly.

On Wednesday (30 March) of this past week Alejandro and I met to discuss the development of a project that (1) has useful outcomes for United Way and for Alejandro's team, (2) requires a realistic amount of effort, given the 10-week term, and (3) has a clearly defined deliverable. We have decided on the construction of one or more simple machine learning classifiers to help predict the outcomes for important benchmarks in United Way's, Break the Cycle of Childhood Poverty campaign.

As part of its commitment to this campaign United Way has awarded **Successful Families 2020** its Community Transformation grant (worth more than $ 3 million dollars over five years). **Successful Families 2020** is a collaborative of six regional nonprofit organizations (Albina Head Start, Immigrant and Refugee Community Organization, Latino Network, Metropolitan Family Service, Native American Youth and Family Center, and Self-Enhancement, Inc.) that serve low-income and culturally specific communities. The goal is to create a Blueprint for Success in Communities of Color and the collaborative has a four-pronged approach to helping these communities. The very first element of this approach is, "Identifying the best-practices that have best served this population through tangible evidence"[1] Alejandro's team, which includes Dr. Ann Curry-Stevens from the School of Social Work here at Portland State, is basically tasked with helping to identify and evaluate this tangible evidence.

I am recommending the use of Naive Bayes classifiers. Alejandro has alerted me to the fact that the **Successful Families 2020** partners have a modest amount of disparate data stored in a wide variety of different database formats. But in many cases there are clear binary outcomes for these data such as "met \ did not meet third-grade benchmarks" or "did \ did not graduate from high school". I recognize that Naive Bayes supervised learning algorithms are relatively decent (but far from perfect!) classifiers and require only a small amount of training data to estimate the

[1] "United Way Announces Strategic Investments to Break the Cycle of Childhood Poverty", 6 June 2014

necessary parameters. And they are fast. I also recognize the independence and prior assumptions that are made by adopting a Naive Bayes classifier of the form,

$$\hat{y} = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i|C_k).$$

Nevertheless, I think such classifiers are potentially useful to Alejandro and his team. As it stands now our project title is (very broadly) something like, "Using Naive Bayes Classifiers to Predict the Probability of Success".

I am very excited about this project! Not only does the project have intrinsic value to United Way, but it is potentially portable to other organizations and bears directly on my own research interests with regards to the feature selection of state variables for evaluating sustainable agriculture at the landscape scale.