JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

12/12/2014

To whom it may concern:

I am writing this reference at the request of Miao Yu. I have interacted with Miao over nine months of courses as one of the professors in the Data Science Specialization offered by the Johns Hopkins University through the MOOC provider Coursera.

The student has taken the following courses through the program:

- **The Data Scientist's Toolbox** (overview of the data, questions, and tools for data analysis, version control software)
- **R Programming** (programming and data analysis in R)
- **Getting and Cleaning Data** (obtaining data from different sources, principles of tidy data and data sharing, data cleaning and manipulation)
- **Exploratory Data Analysis** (basic principles of data graphics, exploratory summaries of data and visualization of multidimensional data using exploratory multivariate statistical techniques)
- **Reproducible Research** (providing reproducibility of data analysis with knitr, R markdown and related tools)
- **Statistical Inference** (fundamentals of statistical inference)
- **Regression Models** (regression models fitting and interpretation, residuals and variability, use of dummy variables and multivariable adjustment, Poisson and logistic regression)
- **Practical Machine Learning** (basic principles of machine learning, building and applying prediction functions with multiple machine learning tools)
- **Developing Data Products** (creating data products using Shiny, R packages, and interactive graphics)
- **Data Science Capstone** (seven weeks long natural language processing capstone project offered in collaboration with SwiftKey)

The program is entirely online; every class includes quizzes, projects and programming.

We acknowledge that MOOC credentials are still being adjudicated in professional and academic circles. However, our program uniquely prepares students to verify their credential. Notably, they will be conversant in the subjects listed above and they will have created deliverables that will be able to highlight their learning process. These include projects such as interactive web pages that call R programs on the back end and prediction algorithms. The student will have a github repository and a codebase of work from the courses. *If this recommendation is for employment purposes, we implore you to give this learner a chance to demonstrate their knowledge and validate their credential in person.*

Miao was one of the first (of around 30) students to complete all nine courses with extra identity verification (called signature track) and pass an invite only tenth capstone project class. This corps of so-called "Dry Runners" were our top students and as such

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University
615 N Wolfe Street, Baltimore, MD 212105

were among 60 who were asked to participate in the dry run.

Being a top student was defined by active and enthusiastic participation in the courses and forums and having performed at the highest level throughout the series. Their participation in the capstone dry run was an invaluable resource in preparing for the over 1,000 students taking the first open run of the capstone available to any student having passed all nine courses signature track.

With over one million enrollments in the program's nine month history, we believe that it is a unique experiment in the history of academics. The Dry Runners played a critical role in its success.

We are committed to this educational model precisely because it enables self learners like Miao. Completion of such a self initiated and paced credential demonstrates dedication, a strong work ethic, and an internal drive for self improvement. Thus, I am delighted to offer this letter of reference.

Sincerely,

Brian Caffo
Professor
Department of Biostatistics
Bloomberg School of Public Health
Johns Hopkins University