The Impact of Curriculum on Graduate Analytics and Data Science Degree Programs

Will Becker, Peggy Tang, An Tiet, Krishna Sonani, Nguyen Ta, Matthew A. Lanham
Purdue University, Krannert School of Management
becke103@purdue.edu; tang371@purdue.edu; atiet@purdue.edu; ksonani@purdue.edu; ta0@purdue.edu; lanhamm@purdue.edu

ABSTRACT

In today's business world, analytics and data science are more important than ever. Universities have developed master's degree programs and courses to satisfy the industry needs. One problem is that data science and analytics often require many different skill sets falling under business communication, quantitative methods, and information systems. In order to estimate the effect that specific courses have on program outcomes, we have developed a predictive model to help students identify which courses to take or help program administrators to modify and decide the curriculum with most impact on student outcomes.

INTRODUCTION

Universities' Data Science or Analytics programs provide students with the proper knowledge to succeed in their field. With a growing demand rate for data scientists/analysts, students will need to know which programs will provide them the best possible curriculum. We will be developing a model by using logistic regression to predict the impact of a program's course offerings on job placement rate.

The below figures show the growing demand for data science, analytics graduates, and some courses that are important to their curriculum. In our study, we plan to provide more quantitative data to analyze the impact these courses have further to supplement the growing demand for data science graduates.

Finding effectiveness of a program's offerings will allow these students to decipher which courses will supplement them best. This study aims to answer the following question:

- What is the impact of the curriculum on a program's effectiveness?

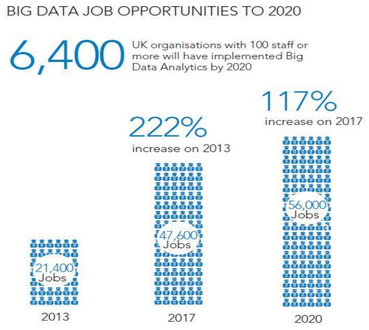

Figure 1: Source

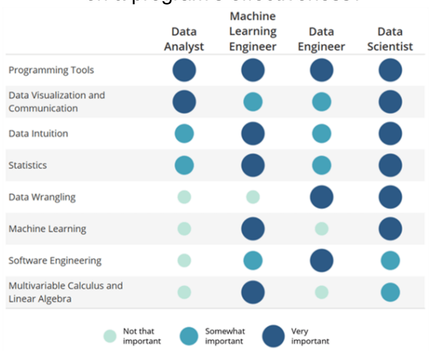

Figure 2: Source

LITERATURE REVIEW

Study	Key Ideas that Assist Our Study
2012, Anthony G. Picciano	Constant change in data science curriculum
2015, Babita Gupta, Michael Goul, Barbara Dinter	Shortage of managers with analytical skills
2015, J. Hardin	Providing the right professors to educate students
2016, II - Yeol Song and Yongjun Zhu	Effects data analytics can bring into a company
2015, Tobias Schoenherr and Cheri Speier - Pero	Steps to take to have a better curriculum

METHODOLOGY

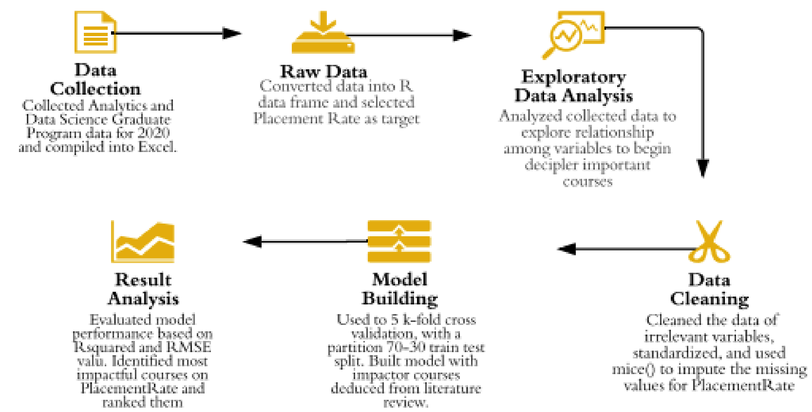

Fig 2. Study Design

STATISTICAL RESULTS

We decided to construct a 5-fold cross-validation model to predict placement rate against the various course offerings we found in our data research.

After running the model, we discovered that our model is unable to effectively predict placement rate against the course offerings per program. Our model achieved an RMSE = 0.0408 and Adjusted R-squared = 0.144, where the RMSE is reasonable and the Adjusted R-squared is not. We believe this issue may arise from the lack of data we were able to gather from website placement statistics.

Figure 3 provides the 2018 placement rate data to compare to the 2020 actual and predicted data that our model discovered in Figure 4.

In Figure 4, we plotted boxplots for our 2020 predicted and actual placement rates. Our predicted values are very inaccurate and have different distribution compared to the actual. We believe this is also due to the low amount of observations that our original data contained.

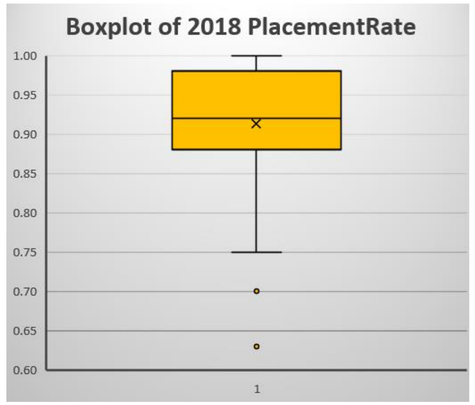

Fig 3. Boxplot of 2018 PlacementRate Data

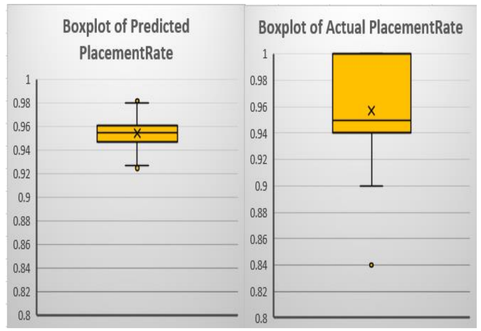

Fig 4. Boxplots of Actual vs. Predicted PlacementRate

EXPECTED BUSINESS IMPACT

Our model will allow students determine which offerings from graduate Analytics and Data Science programs will supplement them the best. It will also allow these programs to determine the best courses to offer to create successful graduates that are placed into jobs after graduation. The results from our model show the overall importance of various courses offered that align with the findings from other studies in our literature review and in Figure 2.

In Figure 5 below, we have ranked the courses that are most effective in predicting placement rate for the Graduate Analytics and Data Science Degree Programs based on their estimated p-value.

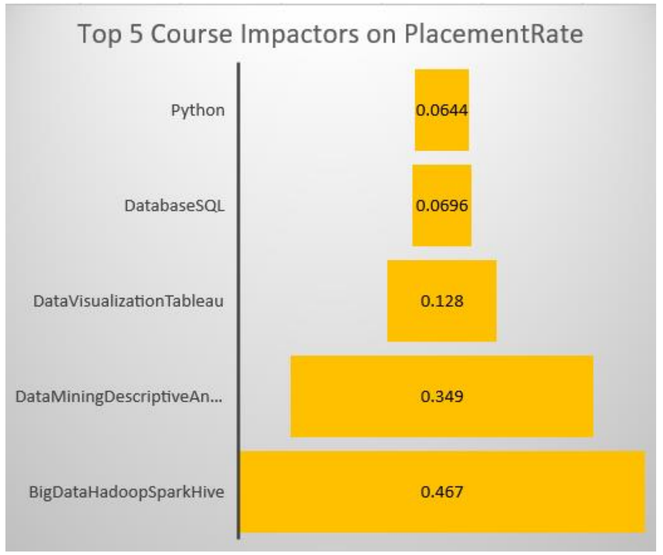

Fig 5. Ranking of Course Prediction Effectiveness

CONCLUSIONS

Overall, our model is not a great predictor of placement rate against the courses offered by each graduate program. Our model does provide some evidence that having **Python**, **SQL**, and **Data Visualization** courses has an association with placement rate.

We believe the main reason our model is a poor predictor of placement is due to the small sample size. Also, incorporating student and program demographic control variables possibly could have improved the model fit.

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham, Xinyu Wang, Theo Ginting, and our graduate student mentors Stefanie Walsh and Hsiao Yu Hsu for their guidance and support on this project.