Does Your Analytics Masters Degree Fit the Bill?
Predicting Placement Rate and Starting Salary

Tyler Richards, Jayana Patel, Matthew Bottorff, Graham Cason, Ersin Nurkic Kacapor, Naveen Ellappan, Matthew A. Lanham

Purdue University Krannert School of Management

richa244@purdue.edu; patel475@purdue.edu; mbottorf@purdue.edu; cason0@purdue.edu; enurkick@purdue.edu; nellappa@purdue.edu; lanhamm@purdue.edu

Abstract

This study builds predictive models that try to estimate two key performance measures (i.e. placement rate, starting salary) of all master's programs in analytics and data science. The motivation for this investigation is that these measures are particularly important to the students in these programs, as well as future students considering which school to attend. Having the ability to identify the key factors that drive program success as we define it, will allow universities to strategically modify their content or program design to recruit better students and improve these KPIs. We collected data from every university and college in America that has a master's program in Analytics or Data Science and built predictive models to help provide insight about what drives success.
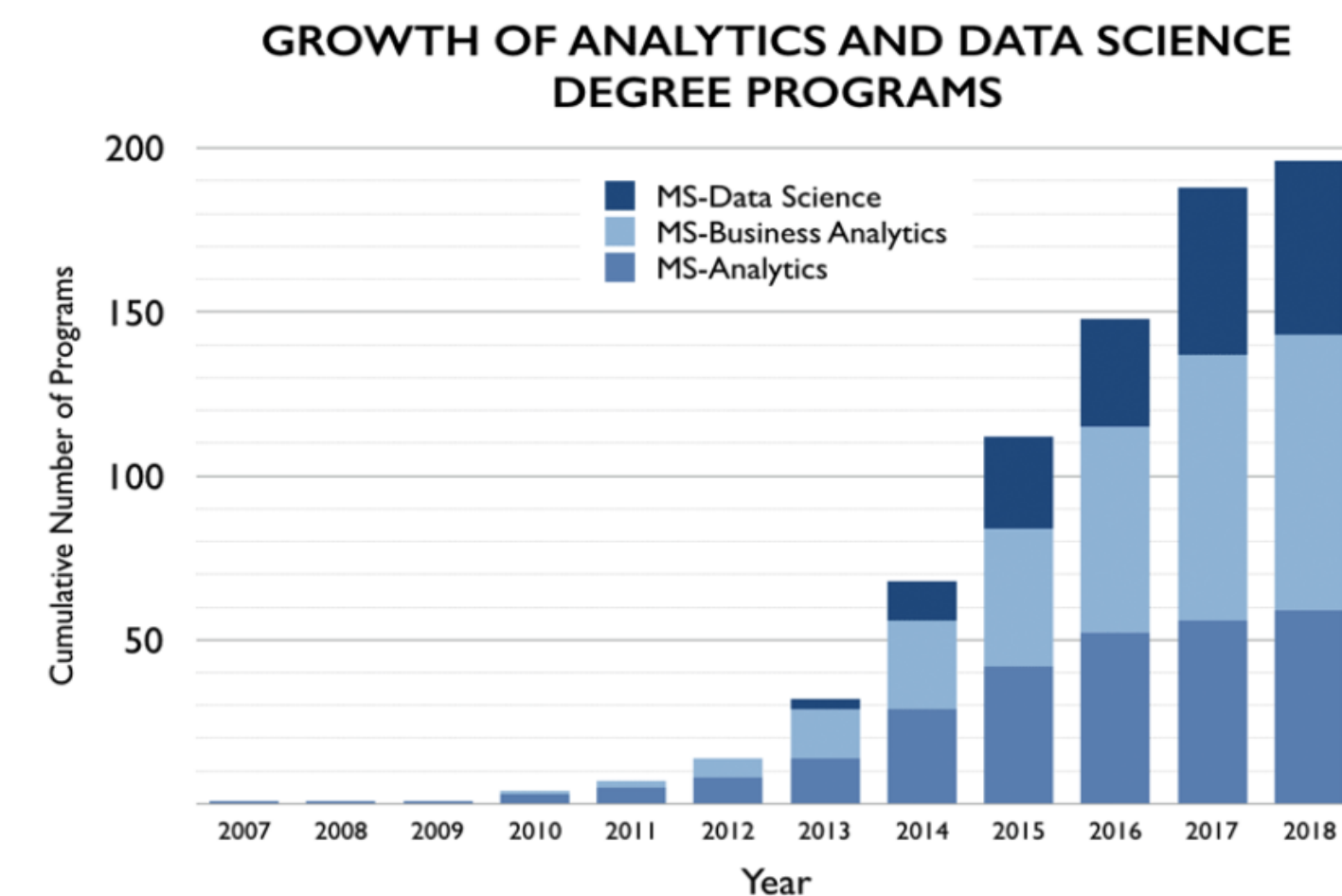
Introduction

Business Analytics (BA has been referred to as the "*Next Frontier for Decision Sciences*" because of the potential impact that turning data into insight can have on a firm's KPIs (Evans, 2012). It has been found that firm's that outperform their competitors are three times more likely to use sophisticated analytics than lower-peered firms (Davenport, 2007). Because industry has placed such importance on analytics, universities have followed suit by modifying their curriculum to offer more data-focused courses, and even creating full masters degree programs in BA and DS.

GROWTH OF ANALYTICS AND DATA SCIENCE DEGREE PROGRAMS

Source:
https://analytics.ncsu.edu

Our research question(s):
➢ What are the key drivers that impact student placement rate and average starting salary?
➢ How well can we predict student placement rate and average starting salary using publicly available information?

Literature Review

After reviewing the literature, we found several articles that discussed BA/DS curriculum suggestions, as well as case studies for teaching "big data" to various audiences. However, we found no published research to date that tried to model "how successful" certain degree programs are based on the students they admit and opportunities they offer. Thus, this study is novel in using predictive analytics to try to improve a program's offering and identifying key drivers that led to stellar program performance, which we define as student placement rate and average starting salary.
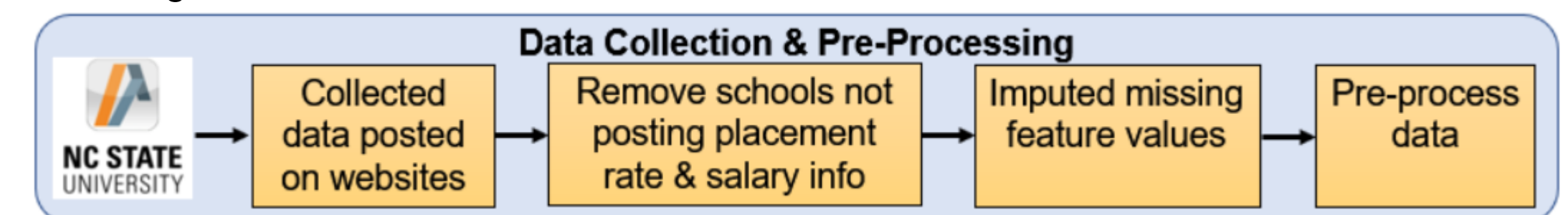
Methodology

Data Collection
The Institute for Advanced Analytics at NC State provided the year the program started and link to their corresponding website. We visited and read about every program and recorded over 50 attributes about them.
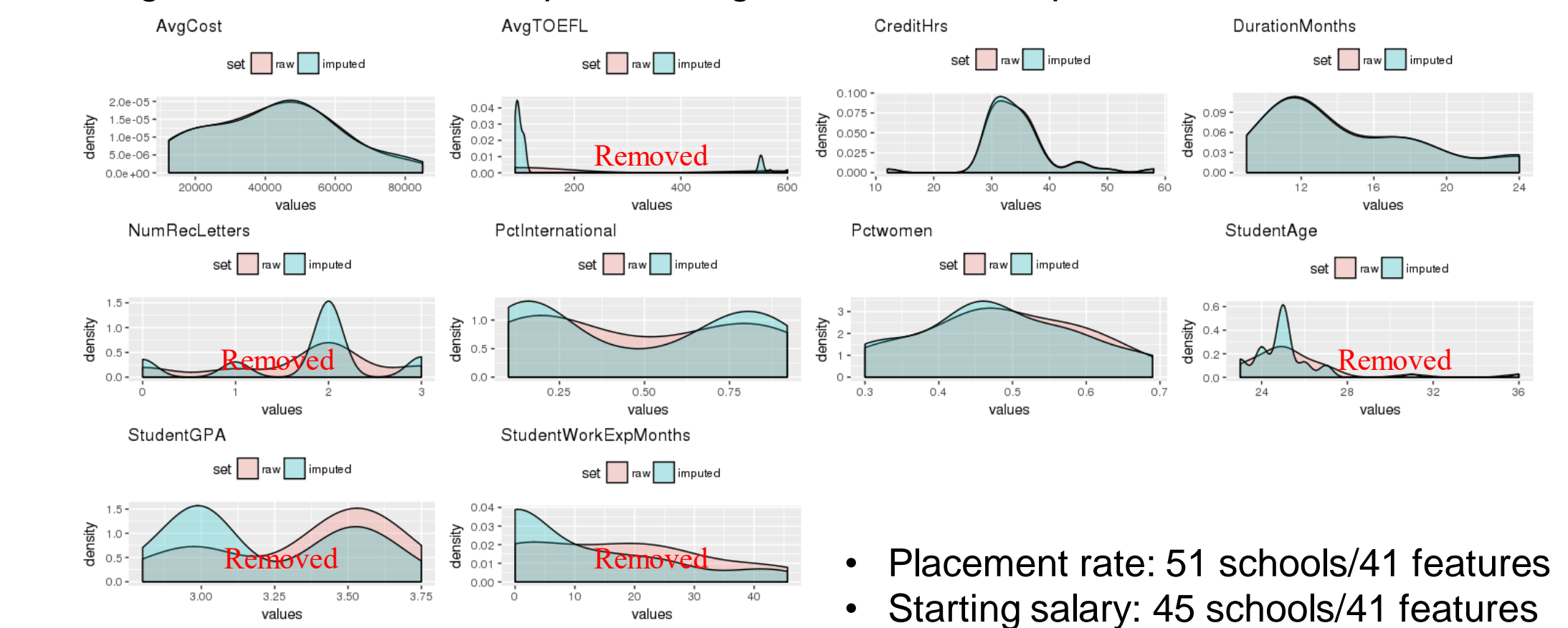
Features Collected

- Program type (online, hybrid, FT/PT)
- Tuition/cost, scholarships
- Student demographics & scores
- Program opportunities (projects, career fairs)
- Courses offered (R, Python, Big Data, etc.)
- Job titles post-graduation
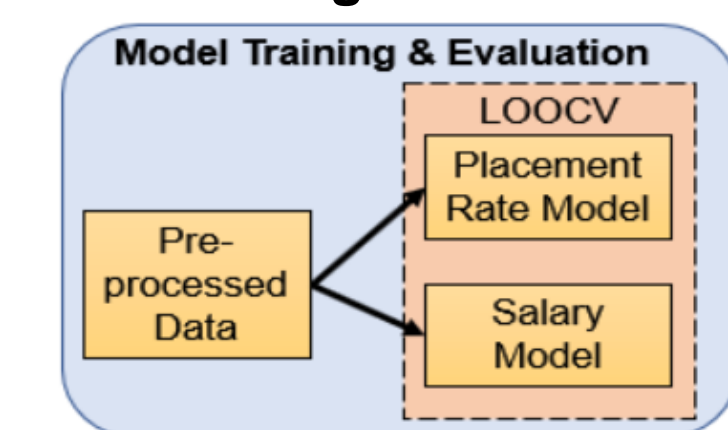- Posted hiring companies

The figure below details how we collected, cleaned, and processed our data for modeling.

Data Collection & Pre-Processing

NC STATE UNIVERSITY → Collected data posted on websites → Remove schools not posting placement rate & salary info → Imputed missing feature values → Pre-process data

- Observations missing placement rate and starting salary were removed
- Program features were imputed using model-based imputation and were evaluated

- Placement rate: 51 schools/41 features
- Starting salary: 45 schools/41 features

Model Design

Model Training & Evaluation

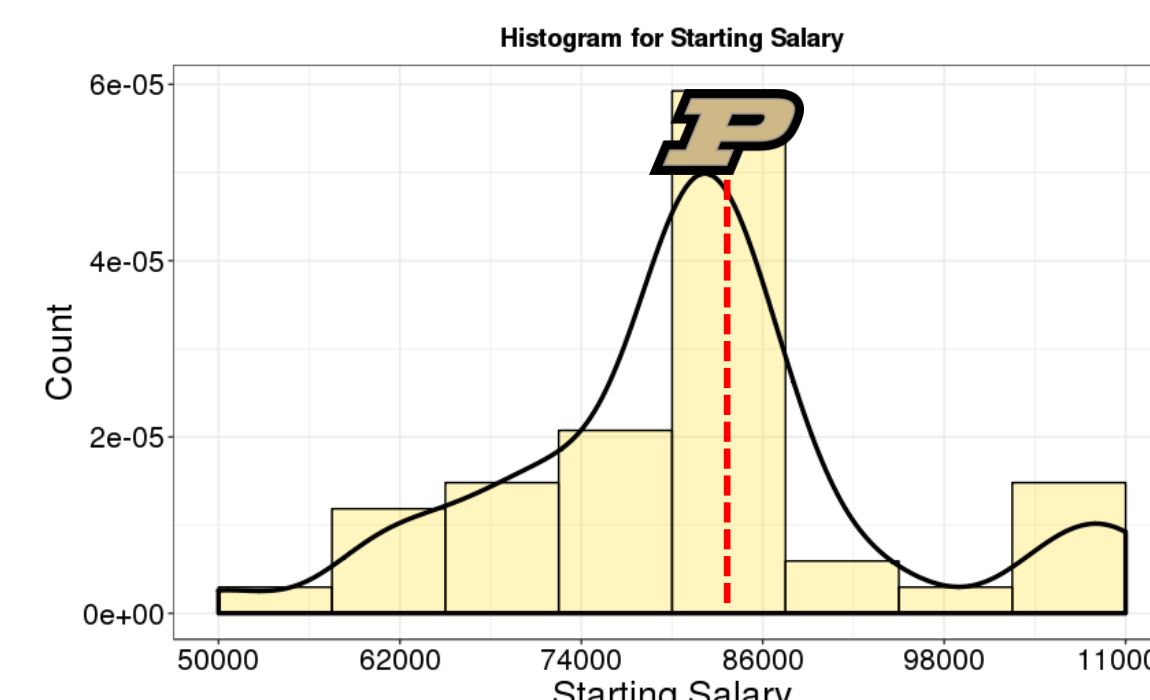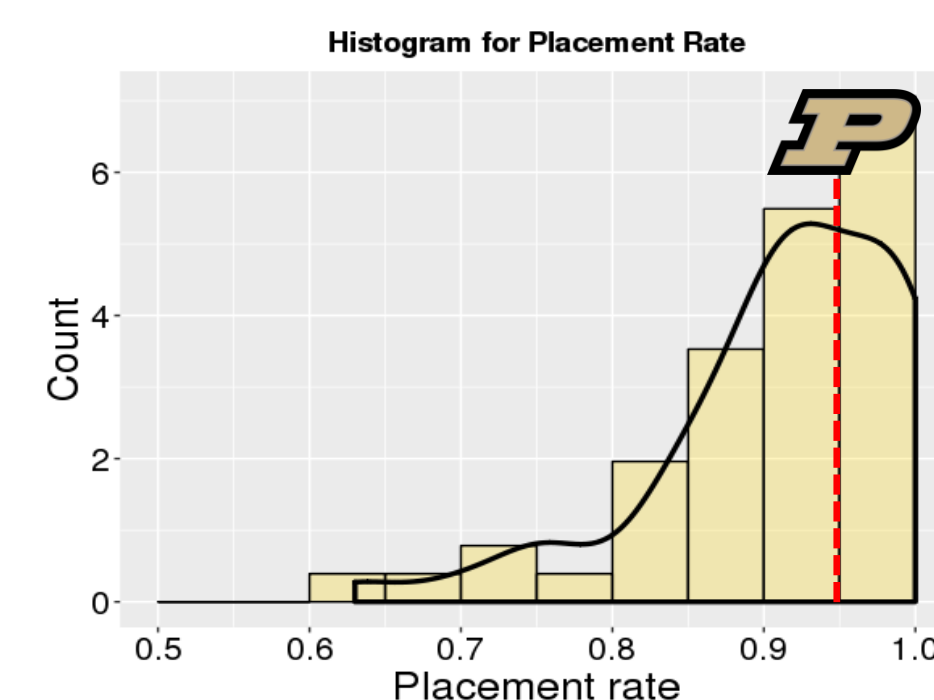Pre-processed Data → LOOCV → Placement Rate Model / Salary Model

Two models were developed; one for placement rate and one for starting salary. Leave-one-out cross-validation (LOOCV) was used assess performance because the data size was small.

Methodology (Approach) Selection
We developed multiple linear regression (forward and backward selection), LASSO, and a random forest (ensemble of trees) models. The justification for these methods is to help us identify important drivers, but also interpret effects with respect to our target variable.
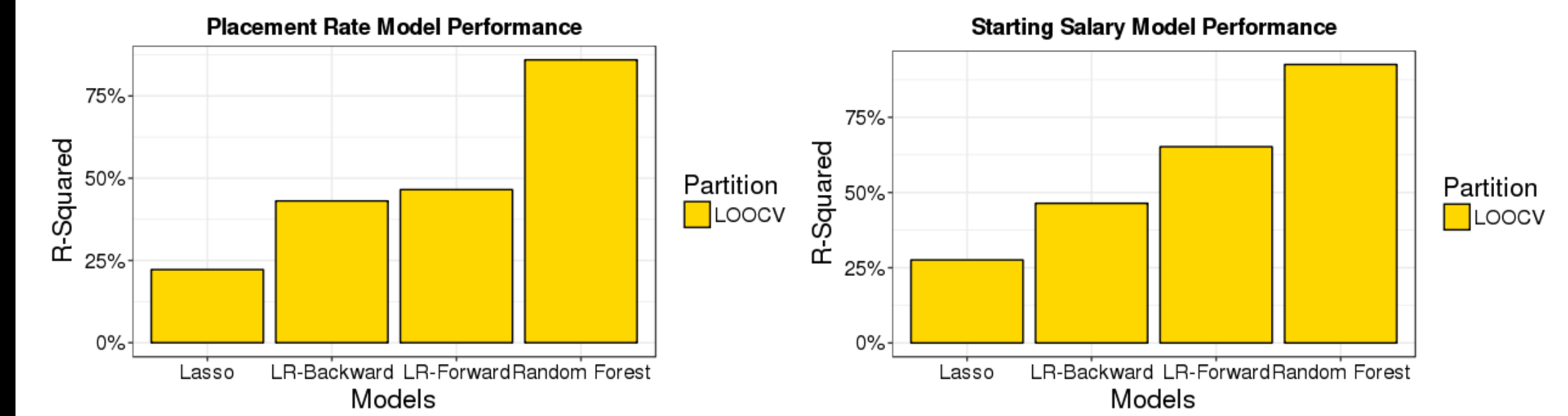
Purdue's M.S. in Business Analytics & Information Management (BAIM) Program
Purdue's MS BAIM program was above average in placement rate at 94% and average (among those that actually report) on starting salary ($81,000).
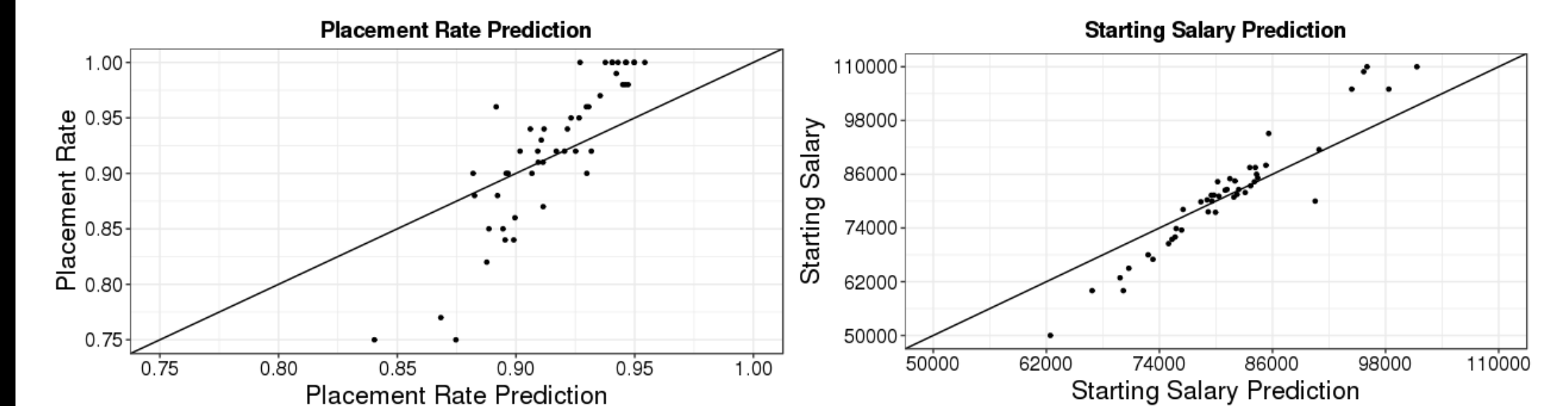
Results

We found that the Random Forest had the lowest error rate (MAE, RMSE) and highest R-squared for both placement rate and starting salary models.

There was more variation in predicting placement rate than starting salary. Most programs will not publish their statistics, which should be a clear indicator to potential students that those programs do not "fit the bill."

Placement Rate Key Drivers

Backward Selection	Forward Selection	LASSO	RF
(-0.16) Analytics	(-0.15) Analytics	(-0.00) Business School	(100) Analytics
(+0.05) Company Sponsors	(+0.06) Business School	(+0.09) STEM Certified *	(73) Data Science
(+0.04) Python	(+0.04) Python	(+0.03) Text Analytics/NLP	(51) Web mining course
(-0.04) SOP	(-0.05) Comm. course	(+0.03) Placement Roles Listed	(49) Career Fairs
(-.014) GMAT/GRE	(+0.08) Scholarships	(-0.01) Interview	(47) Online program
(+0.10) Min GMAT/GRE	(-0.10) GMAT/GRE	(+0.07) Scholarships	(45) SOP
(+0.12) TOEFL scores	(+0.05) Min GMAT/GRE	(+0.05) Min GMAT/GRE	(45) % International
(+0.09) App fee			(45) Database/SQL course
			(44) Placement Roles Listed

Starting Salary Key Drivers

Backward Selection	Forward Selection	LASSO	RF
(-30k) Analytics	(-22k) Years Began	(+21k) Eng. School	(100) Credit Hours
(-24k) Data Science	(+20k) Eng. School	(+18k) Full-time program	(97) Analytics
(-6k) Company Sponsors	(+18k) Full-time program	(+11k) Python	(78) Data Science
(+15k) Data Mining course	(-19k) Analytics	(+10k) Data Mining course	(65) Year Began
(+9k) Text Analytics course	(-19k) STEM Certified	(+3k) Data Vis course	(59) Comm. Course
(-7k) R course	(+5k) Python	(+5k) Big Data course	(51) % International
(-15k) Credit Hours	(+24k) TOEFL scores	(-2k) Text Analytics course	(45) Company Sponsors
(+18k) GMAT/GRE	(-14k) Scholarships	(+8k) GMAT/GRE	(43) GMAT/GRE
		(+2k) TOEFL scores	(39) Data Mining course
		(+2k) Placement Roles Listed	(36) Pred. Analytics course

Program | Easily Changed | Administrative

Conclusions

Our study found that we could predict a program's placement rate and starting salary, but there are other important casual factors we are not measuring that is impacting placement rate. A program could improve their performance in these areas if they focus on decisions they have control over, such as course offerings, and modifying administrative/admission's tasks.

Acknowledgements