# **Will They Come Back?**

# **– A Predictive and Interpretable Model for Student Retention at UNC**

Consultants: xxxxxx

Client: Lynn Williford, Institutional Research and Assessment, UNC at Chapel Hill

Professor: Dr. Perry Haaland and Dr. Steve Marron

January 15, 2020

**Abstract**

This report created a model and framework to proactively identify students who are at risk of not returning in the next semester. Using Generalized Linear Models and Decision Trees, we identified several key factors that significantly impact the probability for a student to return to school in the next semester. In general, the probability of returning is higher for a student with more credits earned, with less credits failed, having a double major status, being a new student (instead of a transfer student), being in certain programs (such as business), being a junior, and being an in-state student. This would potentially help school officials to target and engage with students at risk early on in order to provide guidance and assistance to not only benefit the students but also enhance the University’s reputation.

**Background**

(Current status, Need, Significance of this work, Target audience)

**Data**

(Source, number of samples, variables, descriptive figures if needed)

**Method**

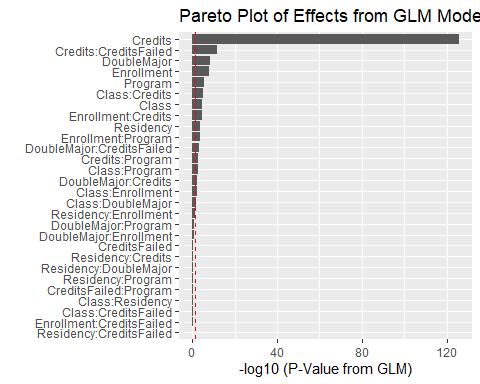
(GLM, Decision tree, others?)

**Results**

1. *Identification of key factors for student retention -- Angel*

Xxx (GLM model, significant variable, p-value, etc.)

Figure 1 Ranking of key effects by significance from GLM Model



*Note.* The Pareto plot above shows the terms from the GLM logistic regression fit with Retention as the response. The terms are sorted from top to bottom from most to least significant. The model contained the main effects and all of the possible two-factor interactions: Credits, CreditsFailed, DoubleMajor, Enrollment, Program, Class, Residency. Red dotted line indicates where p-value = 0.05. The bars that exceeds the red dotted line imply that the corresponding effect is statistically significant with α = 0.05.

1. *(Credits and CreditsFailed) – Angel*
2. *(DoubleMajor) – Zhaoqi & Angel*
3. *(Enrollment) -- Xinjie*
4. *(Program) – Angel & Zhaoqi*
5. *(Class) – Qinghua & Ruiping*
6. *(Residency) – Ruiping & Qinghua*

**Conclusion**

(1 sentence summary from each result)

(Weakness/Future directions, eg. including GPA, financial aid information, etc.)