Machine Learning Assisted College Advising:

A Data Based Exploration of College Success

Eon Slemp

Every year millions of high school students enroll in a 4 year college.

- About a 40% will not graduate
- average of $13,000 of debt
- 50% will default on this debt
- DIfficult to quantify psychological effects

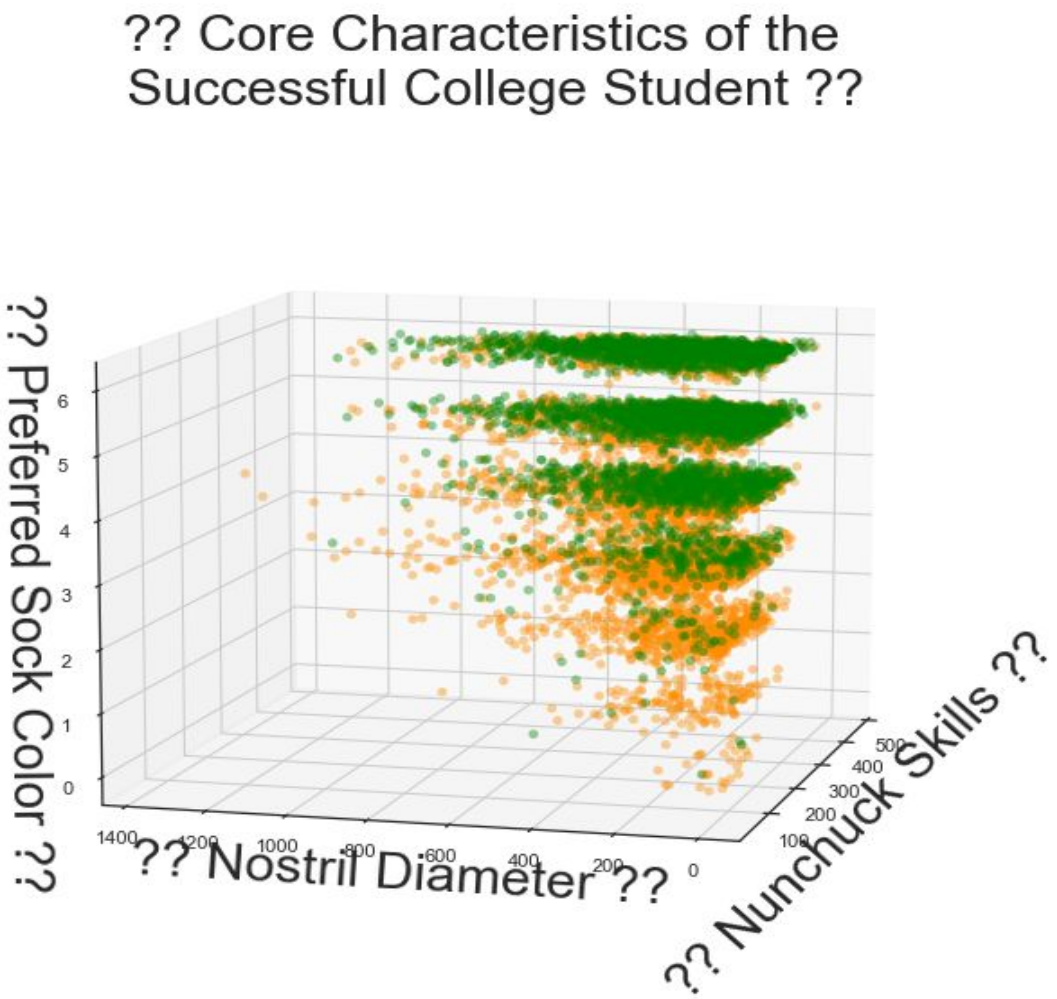
Objectives of this analysis

1. Identify traits of successful college students
2. Develop a machine learning model to predict college graduation

What are the core attributes of a college graduate?

- Nostril diameter?
- Nunchuck Skills?
- Prefered sock color?

What is the nature of this beast?



Educational Longitudinal Study of 2002

- Starts in sophomore year of high school
- 3 follow up collection points over next decade
- Over 16,000 participants
- Over 5000 features
- About 5% of data NaN values

Data used in this analysis :

- Selected 31 features
- 9900 rows
- Imputed missing values - KNN imputation
- Extensive feature engineering
- High Interpretability

Feature Categories

Home Life/ lifestyle:

- Risk Factors
- Literacy resources
- Screentime
- Work
- Aspirations
- Xtra-curriculars
- Sports
- Aspirations

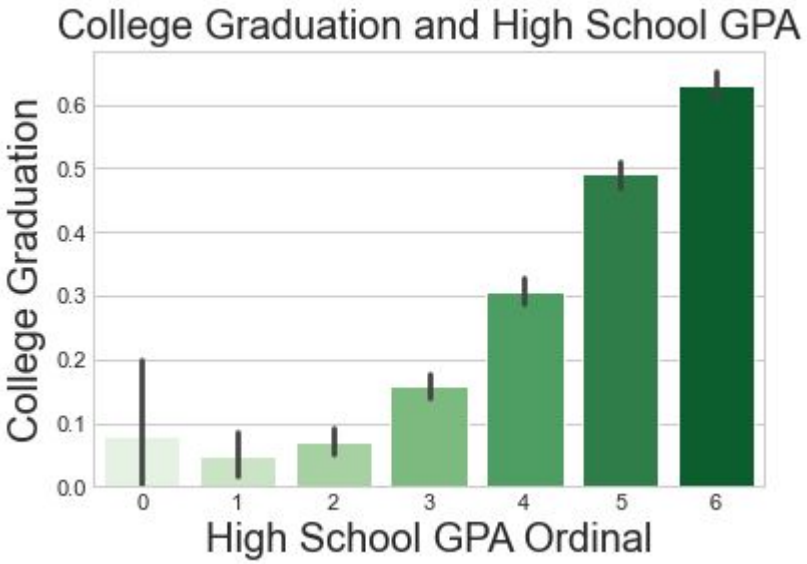
Academics:

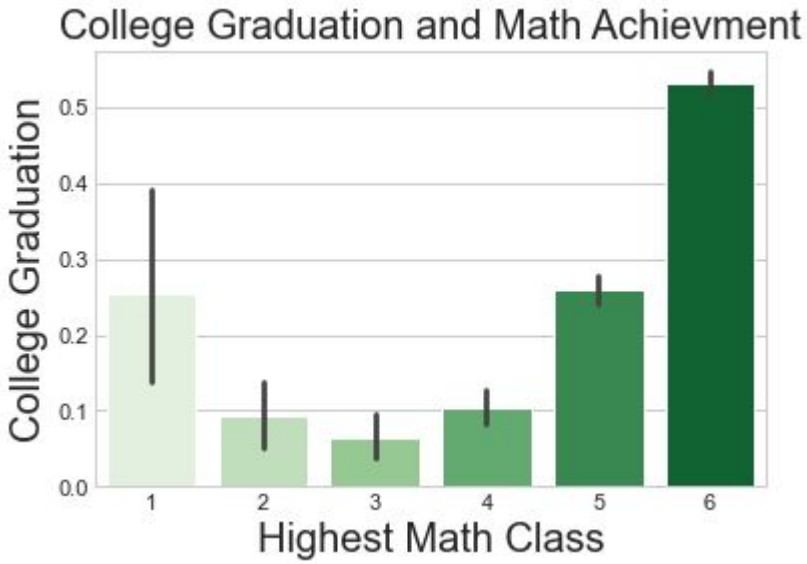
- Standardized Testing
- High school GPA
- Academic load
- Highest math class completed
- Educational confidence

Engineered:

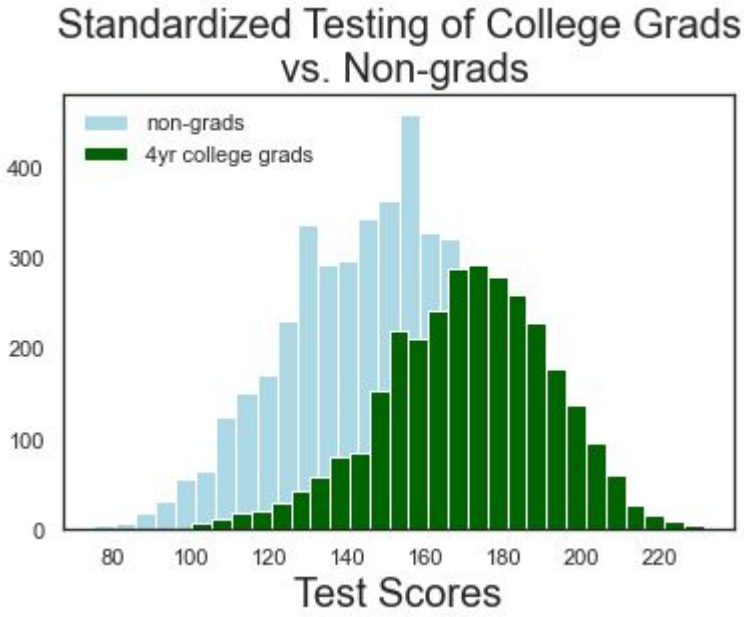
- Verbal ability
- Mathematical ability
- Concientiousness
- Effort
- Academic performance
- Delinquency

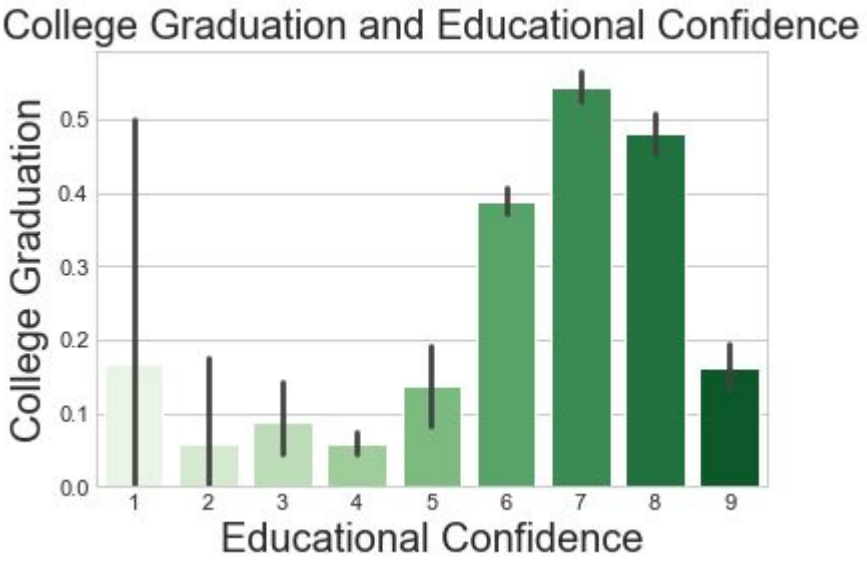
The Traits of College Graduates



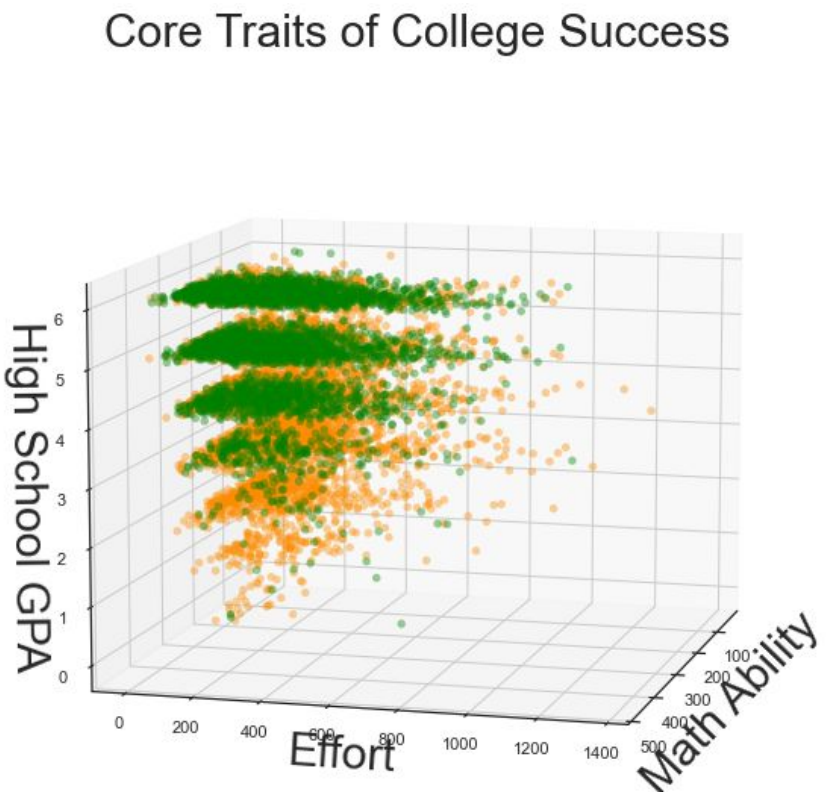


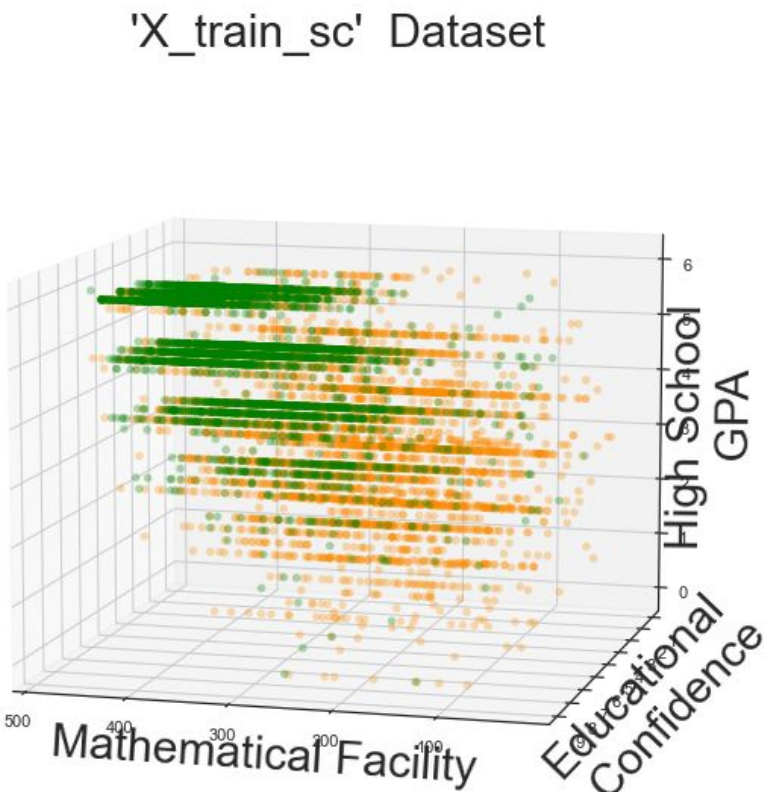
The Traits of College Graduates





College Success





Modeling

Algorithms:

- Random Forest
- Logistic Regression
- XGBoost
- Gaussian Naive Bayes
- KNN

Data Sets:

- 64 features with dummy variables
- SelectKBest 25, 20, and 15 features - scaled & unscaled
- Scaled 3 feature set
- Target = graduation from 4 yr college

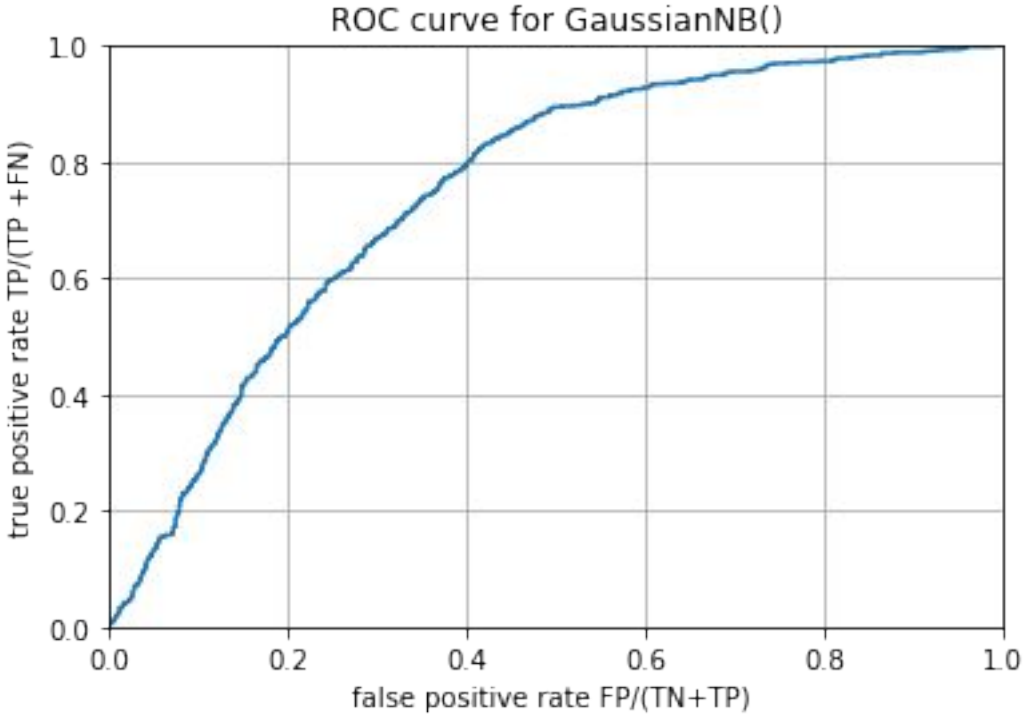
Best Model : Gaussian Naive Bayes

Accuracy = 0.6835

F1 = 0.6511

Recall = 0.7405

Precision = 0.5809



Application:

- Results are a data point not an answer.

ROI and Recall