Predictors of High School Dropout
Beyond Risk Factors

Salesforce Data Science Challenge:

Tawny Tsang
Landon Gibson
Arad Saebi
Lilian Jiang
Terry Allen

UCLA

Do we know whom we serve?

The dropout challenge

74% of high school dropouts regret their decision (Bridgeland et al, 2005)

Students cite inadequate social, emotional, and academic support (Bridgeland, 2010)

Increasing support and guidance for post-secondary options reduce dropout rates by 10%, but are costly (~$14,000/student) (Pathways—Oreopoulos et al., 2014)

Predictive Analysis: Results

What do our data reveal?

Model Performance on Test Data

Model Description	AUC	AUC: 95% Lower CI	AUC: 95% Upper CI	Accuracy	Specificity	Sensitivity
Naïve Bayes	0.978	0.954	1.000	0.946	0.903	0.967
Random Forest	0.964	0.929	0.999	0.891	0.742	0.967
Gradient Boosted Machine	0.961	0.926	0.997	0.913	0.903	0.918
Boosted Classification Tree	0.959	0.923	0.994	0.902	0.871	0.918
Elastic-Net Regularized General Linear Models	0.958	0.923	0.993	0.902	0.903	0.902
Neural Network	0.938	0.890	0.986	0.880	0.903	0.869
AdaBoost Classification Trees	0.907	0.839	0.974	0.891	0.839	0.918

Note: Positive case is when students do not dropout

A Closer Look at Test Predictions

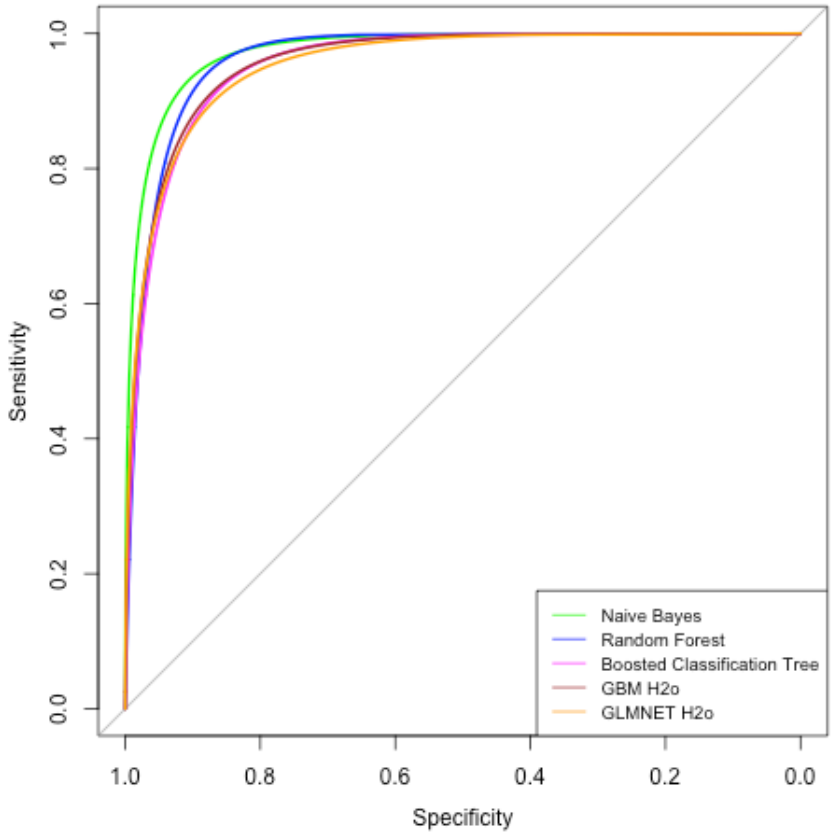
Naive Bayes

	Truth	
Prediction	Didn't Dropout	Dropped Out
Didn't Dropout	59	3
Dropped Out	2	28

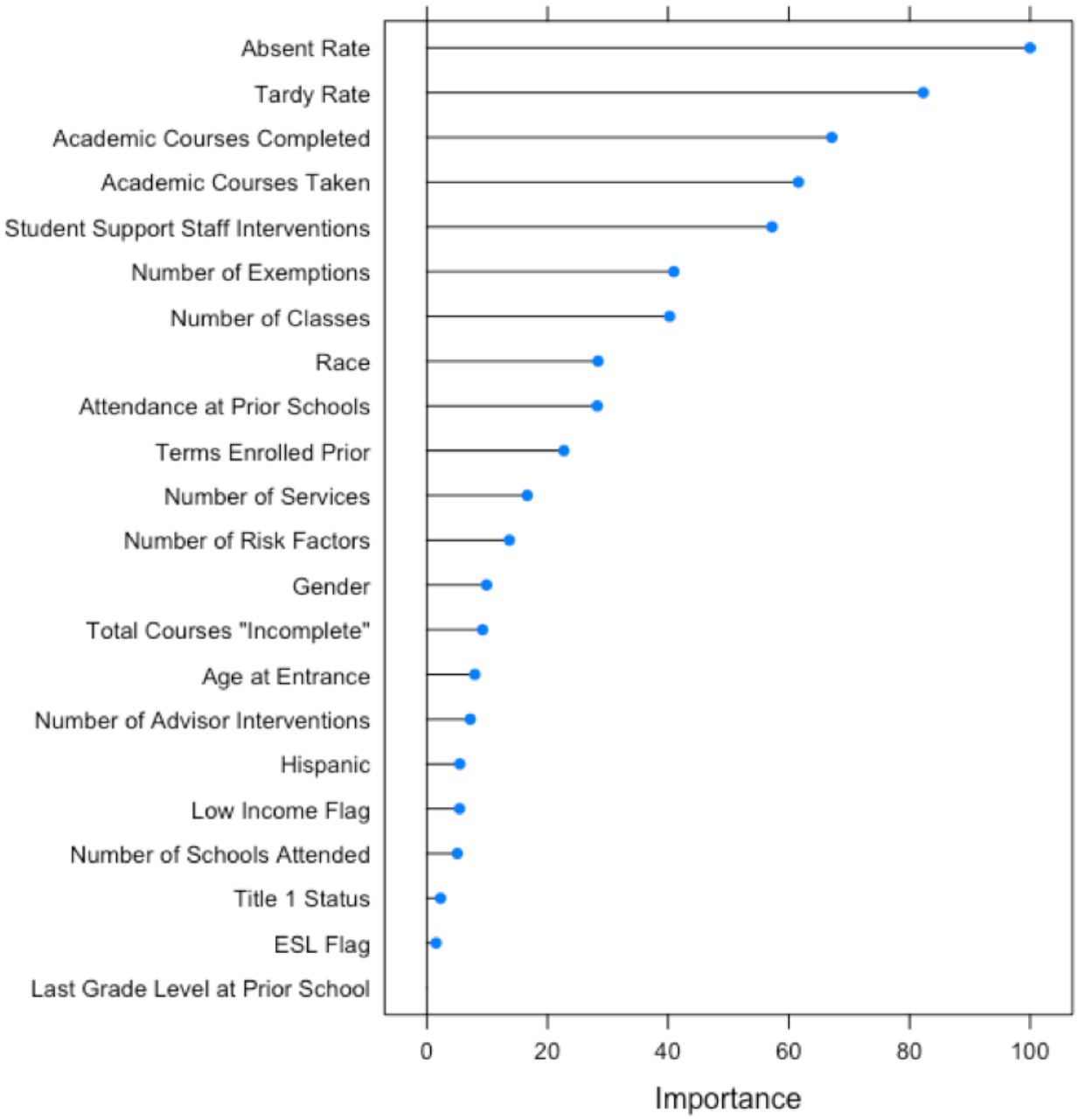
Specificity = 28 / (28 + 3)

Sensitivity = 59 / (59 + 2)

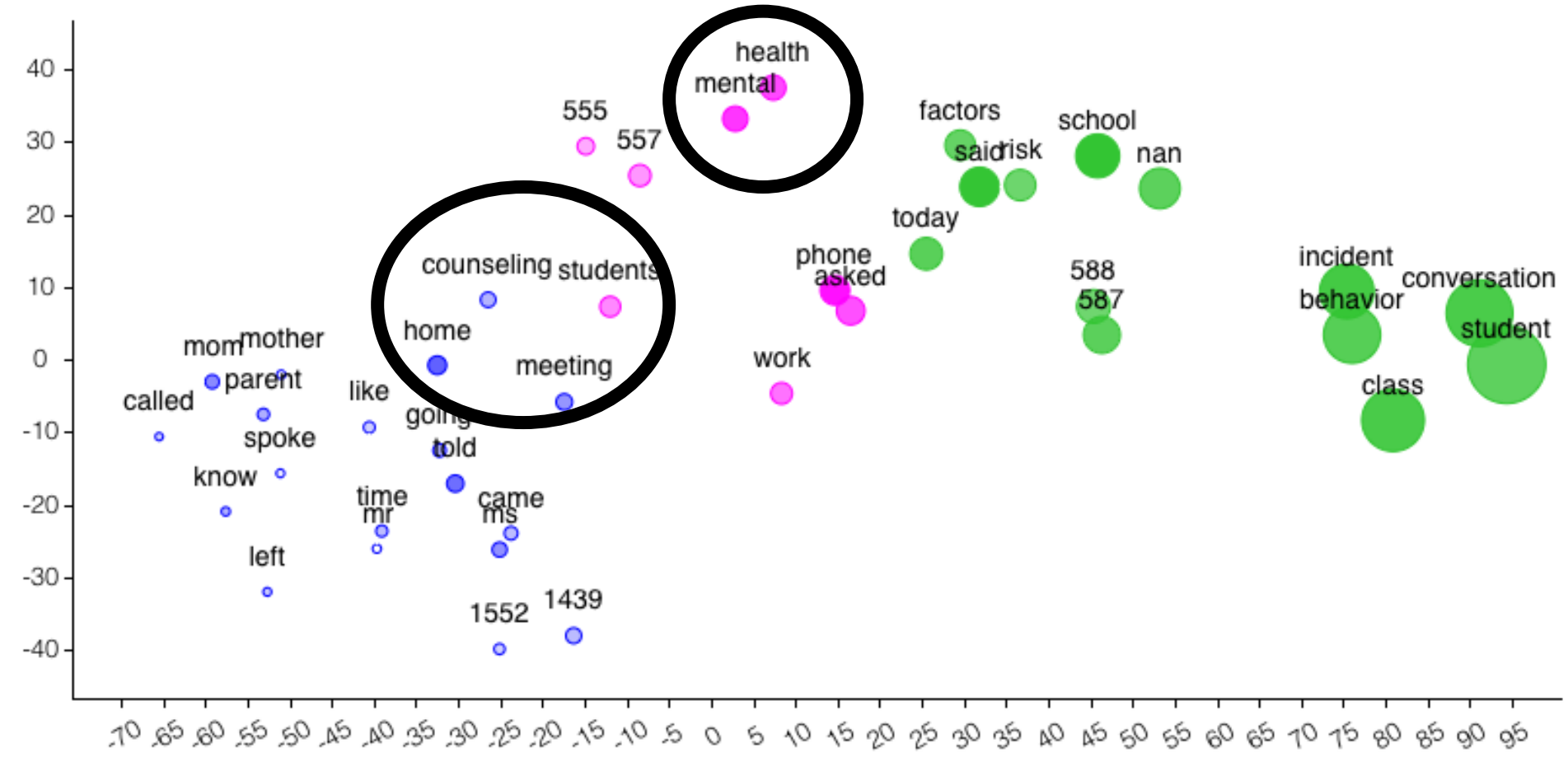
Graphical Comparison



Naive Bayes Model predicts which students will dropout with 90.3% Specificity and 96.7% Sensitivity on test data


Absent Rate
Tardy Rate
Academic Courses Completed
Academic Courses Taken
Student Support Staff Interventions
Number of Exemptions
Number of Classes
Race
Attendance at Prior Schools
Terms Enrolled Prior
Number of Services
Number of Risk Factors
Gender
Total Courses "Incomplete"
Age at Entrance
Number of Advisor Interventions
Hispanic
Low Income Flag
Number of Schools Attended
Title 1 Status
ESL Flag
Last Grade Level at Prior School
0
20
40
60
80
100
Importance

Textual analysis



Conclusions

Student involvement in educational plan may increase motivation and interest in schooling

For nontraditional students, the importance of long-term benefits need to be apparent to offset short-term costs of completing secondary school

Dropping out of school is related to a variety of factors that can be classified in four domains: individual, family, school and community factors

Future Directions

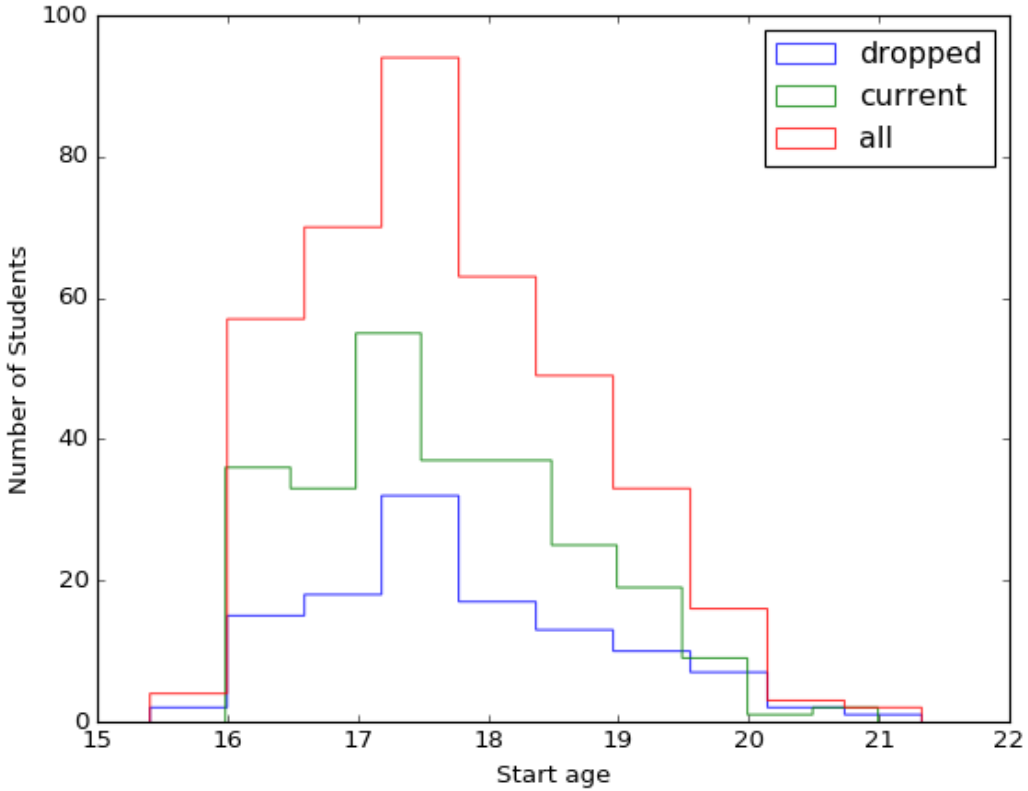
Researchers

- More preprocessing & tuning
- Utilizing text data in predictive models
- Capturing all four domains: individual, family, school and community factors

Practitioners

- Identify key transition points in students' school career
- Build an Early Warning System
- Implement culturally relevant and responsive curriculum and instruction
- Use a Tiered Prevention and Intervention System grounded in personalization, care, the transition to "adulthood"

Age (??)


dropped
current
all
Number of Students
Start age
0
20
40
60
80
100
15
16
17
18
19
20
21
22

Factors Associated with Dropout

Generating prediction model of dropout rates in BDEA students

What were the covariates?

Analyze transcripts between students and counselors