Sam Houston State University

February 2025

Predicting First-Term Academic Standing for First-Time Freshmen

Kristina Beltran, Heather Caudle, Brandon Cooper, Glenn Harris, Nicholas Hudzinski, and Tristan Young

tay011@shsu.edu & neh007@shsu.edu

OUTLINE

1

Background

- Current intervention methods
- Multi-department initiatives
- Brainstorming new early intervention methods

2

Methods

- Random Forest (RF) Modeling
- Linear Regression

3

Data and Model

- Developing a predictive and inferential tool
- Tool deployment in progress reporting periods

4

Discussion

- Model disagreement
- Post-term assessment and findings (Fall 2024)

BACKGROUND

Institutional Goal:

Improve 1-semester persistence for first-time freshmen (FTF) by identifying at-risk students and directing intervention campaigns towards these students.

Early Alert Gaps:

Looking at our Fall 2023 cohort, 48.7% of at-risk FTF had no contact with academic support services.

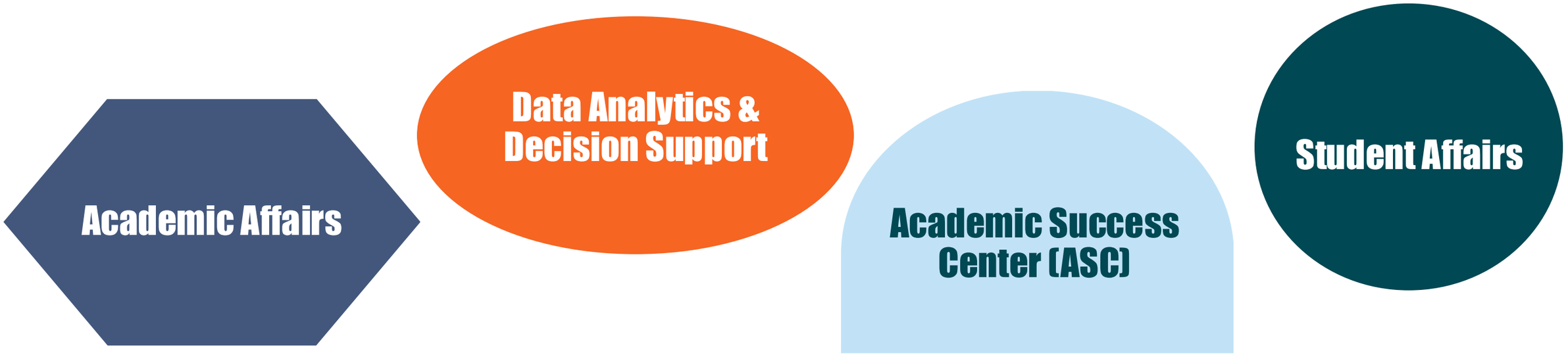
Develop a model to identify students missed by academic support services early in the semester.

Provide a tool to help them contact these additional students.

Early Intervention Needs

- Importance of identifying at-risk students as early as possible
- Current progress reporting system flags students in weeks 4 and 7
- Need for predictive modeling to supplement early alerts

CROSS-FUNCTIONAL COLLABORATIONS



OBJECTIVE: Develop a predictive model and intervention tool

GOAL

Improve FTF first-term academic outcomes by creating a tool using predictive and inferential models to better target students at risk

What factors put students at risk?

↓

Linear Regression Model (LM)

Inferential Results

↓

New Tool Development & Implementation

What is the likelihood of ending term in good standing?

↓

Random Forest Model (RF)

Predictive Results

↓

Validation

METHODS

Classification Problem

→ Predicting good vs. unsatisfactory standing

Linear Regression (LM) Model

→ Predicts first-term GPA.

→ Predicted GPA < 2.00 would be unsatisfactory academic standing.

Utilization of Random Forest (RF) Model

→ Predicts academic standing as a binary outcome

→ Used to check for agreement with LM

Data, Handling, & Analysis

Dataset: 11,776 FTF records (Fall 2021-Fall 2024) considered. 170+ variables

Feature Selection: LASSO regression for RF, stepwise selection for LM (BIC)

Balancing Strategy: Oversampling of at-risk students for RF training

Missing Data Handling: Indicators for missing values + identical imputation methods used for both RF and LM

VARIABLES CONSIDERED

Demographics

Gender

First-generation status

Ethnicity

Campus Employment

Athlete

Academic Indicators

High School GPA

SAT scores

TSI readiness

Dual Credit Indicator

Feeder School Indicator

College

Department

Financial Aid

Pell grant status

State or Institutional aid

Student Behavior

Blackboard course logins

Attempted assignment rates

RESULTS

Variable	Estimate	Std.Error	t.value	p.value
(Intercept)	-1.21	0.13	-9.51	0.00
genderMale	-0.08	0.02	-4.08	0.00
first_generation_statusNot First Generation	0.13	0.02	6.14	0.00
first_generation_statusUnknown	0.06	0.03	2.12	0.03
collegeBusiness Administration	-0.24	0.04	-6.47	0.00
collegeCriminal Justice	-0.15	0.04	-4.24	0.00
collegeEducation	-0.15	0.05	-3.26	0.00
collegeHealth Sciences	-0.24	0.04	-6.46	0.00
collegeHumanities and Social Sciences	-0.22	0.04	-5.31	0.00
collegeScience and Engineering Technology	-0.34	0.03	-10.12	0.00

Variable	Estimate	Std.Error	t.value	p.value
tsi_math_ready_censusY	0.21	0.02	8.51	0.00
tsi_rdng_ready_censusY	0.23	0.03	8.66	0.00
high_school_gpa	0.59	0.02	25.66	0.00
sat_concordance	0.00	0.00	10.10	0.00
sat_concordance_missing	-0.15	0.02	-7.51	0.00
received_pellY	-0.19	0.03	-5.93	0.00
state_aid_no_loans_work	0.00	0.00	4.52	0.00
private_aid_no_loans	0.00	0.00	5.19	0.00
gateway_math_complete_cY	0.17	0.03	6.71	0.00
athlete_indY	0.49	0.06	8.28	0.00
low_log	-1.03	0.05	-21.75	0.00
high_log	0.65	0.04	15.23	0.00
assign_ratio	0.52	0.04	14.33	0.00
asc_appt_count	0.07	0.01	10.07	0.00

RESULTS

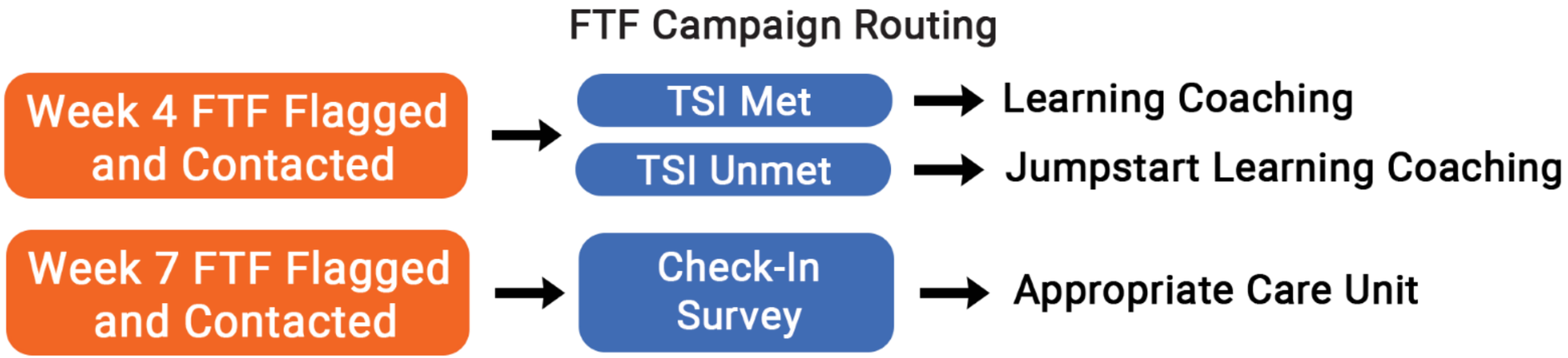
LM Model Performance

- **Four-Week Model Accuracy:** 77.52%
- **Seven-Week Model Accuracy :** 77.73%
- **True Negative Prediction Rate for students with no ASC intervention (Seven-Week Model):** 56.58%
- **Significant Variables:** High school GPA, TSI readiness, Blackboard engagement, Demographics

RF Model Performance

- **Four-Week Model Accuracy:** 78.2% overall
- **Seven-Week Model Accuracy:** 77.92% overall
- **True Negative Prediction Rate for students with no ASC intervention (Seven-Week Model):** 53.70%
- **Key Predictors:** High school GPA, TSI readiness, Blackboard engagement, Demographics

Tool Deployment (Academic Success Center)



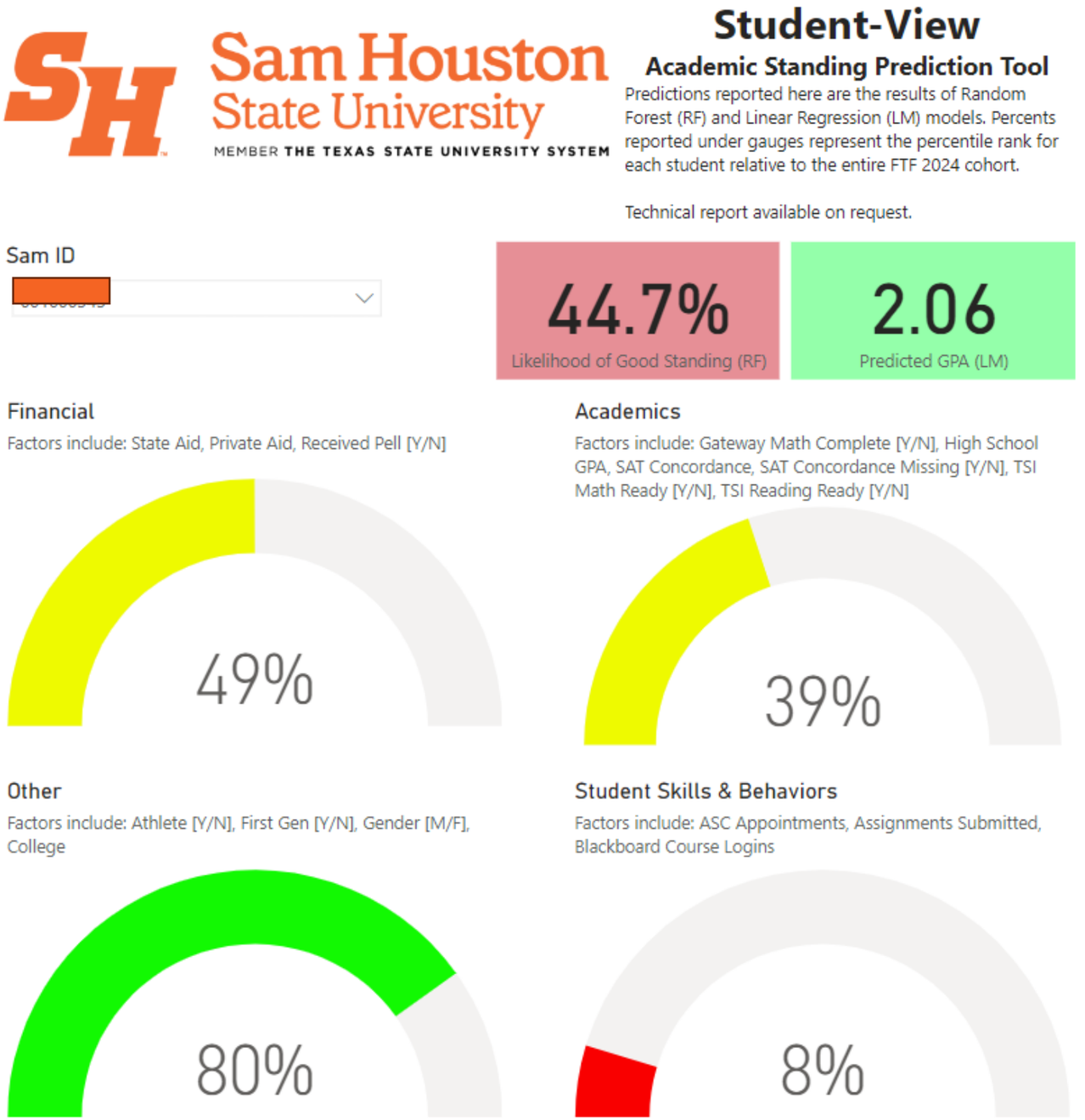
Purpose: Assist ASC staff in targeting interventions.

Student View: Individual risk factors visualized for advisors for a single student.

List View: Provides risk status of all students in exportable table format.

ACADEMIC STANDING PREDICTIVE TOOL

STUDENT VIEW



ACADEMIC STANDING PREDICTIVE TOOL

LIST VIEW

Sam Houston State University
MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

List-View

Academic Standing Prediction Tool

Predictions reported here are the results of Random Forest (RF) and Linear Regression (LM) models. Table results can be exported by hovering over the table and clicking on the ellipses representing "More Options."

Technical report available on request.

Sam ID: All

Good Standing (Hold CTRL to Multiselect): All

Alerted?: All

Campaigned?: All

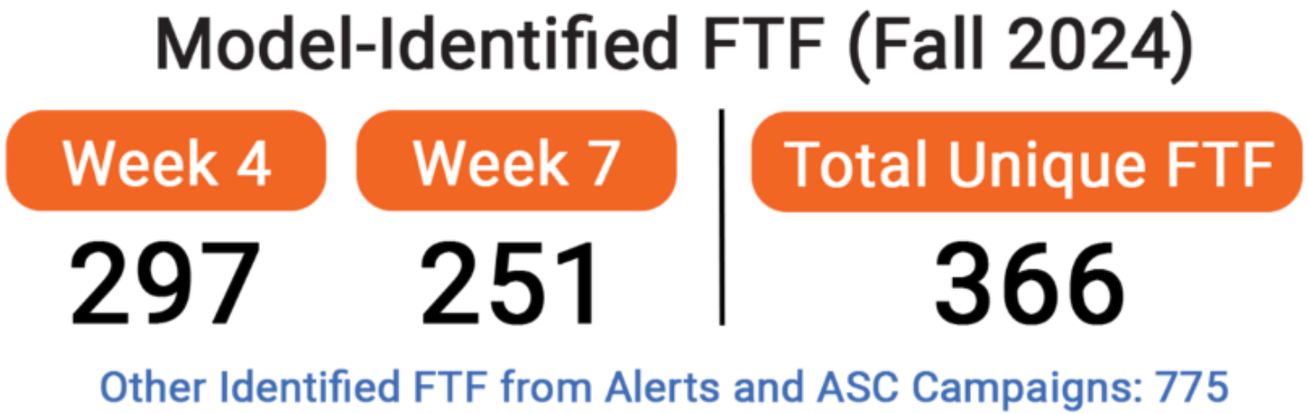
College: All

Department: All

3274
Total Students

Sam ID	College	Department	Good Standing (RF)	Likelihood of Good Standing (RF)	Good Standing (LM)	Predicted GPA (LM)
[redacted]	Humanities and Social Sciences	Political Science	Y	61.7%	N	1.74
	Arts and Media	Art	Y	50.5%	Y	2.09
	Science and Engineering Tech	Agricultural Sciences	Y	54.4%	N	1.85
	Science and Engineering Tech	Biological Sciences	Y	67.3%	N	1.96
	Criminal Justice	Criminal Justice & Criminology	N	39.6%	N	0.97
	Science and Engineering Tech	Biological Sciences	N	34.1%	N	0.99
	Business Administration	Finance & Banking	N	31.8%	N	1.15
	Science and Engineering Tech	Agricultural Sciences	N	43.3%	N	1.71

Current Insights



1141

Total unique FTF intervened by the ASC thus far (including model-identified students)

Current Impact

366 additional FTF reached through predictive models. 80 of the 297 students identified by the models/tool in week four were later reported by their faculty in progress reports in week seven

Future Directions

- Incorporate feedback from ASC regarding the tool
- Adapt model to predict Spring academic standing for students that persisted from Fall
- Add new variables such as a weighted DFQ rate and IDEA data

CONCLUSION:

- Predictive modeling enhances early intervention efforts.
- Institutional support for at-risk students can be improved through data-driven decision-making.
- Future iterations and ASC feedback will refine the model and intervention strategies.

Discussion

QUESTIONS?

Scan the QR code to complete the session survey.

Texas Association for Institutional Research

Annual Conference: February 25-28, 2025
Omni Hotel in Corpus Christi, TX