Retention and Student Success

Beyond Retention - An Integrated Student Success Model

Prabin Raj Shrestha '24 Hope Smalling, Ed.S.

Charge:
Engage students in assessment process

Timeline:
Academic year - Fall 2023 through Spring 2024.

Role:
Support creation of reports from early alert data. Use best practices to analyze large data sets.

Goals
Enhance the assessment and use of the early alert system (EAS) and related outcomes

1. **Better understand the relation between the selected flags and grade outcome**
2. Qualitative analysis of notes data from flags and kudos

Questions
- How will notes entered by faculty and advisors support improvement of retention and student success?
- Can notes data be used as a leading indicator of student success?

Assessment of Future Practice

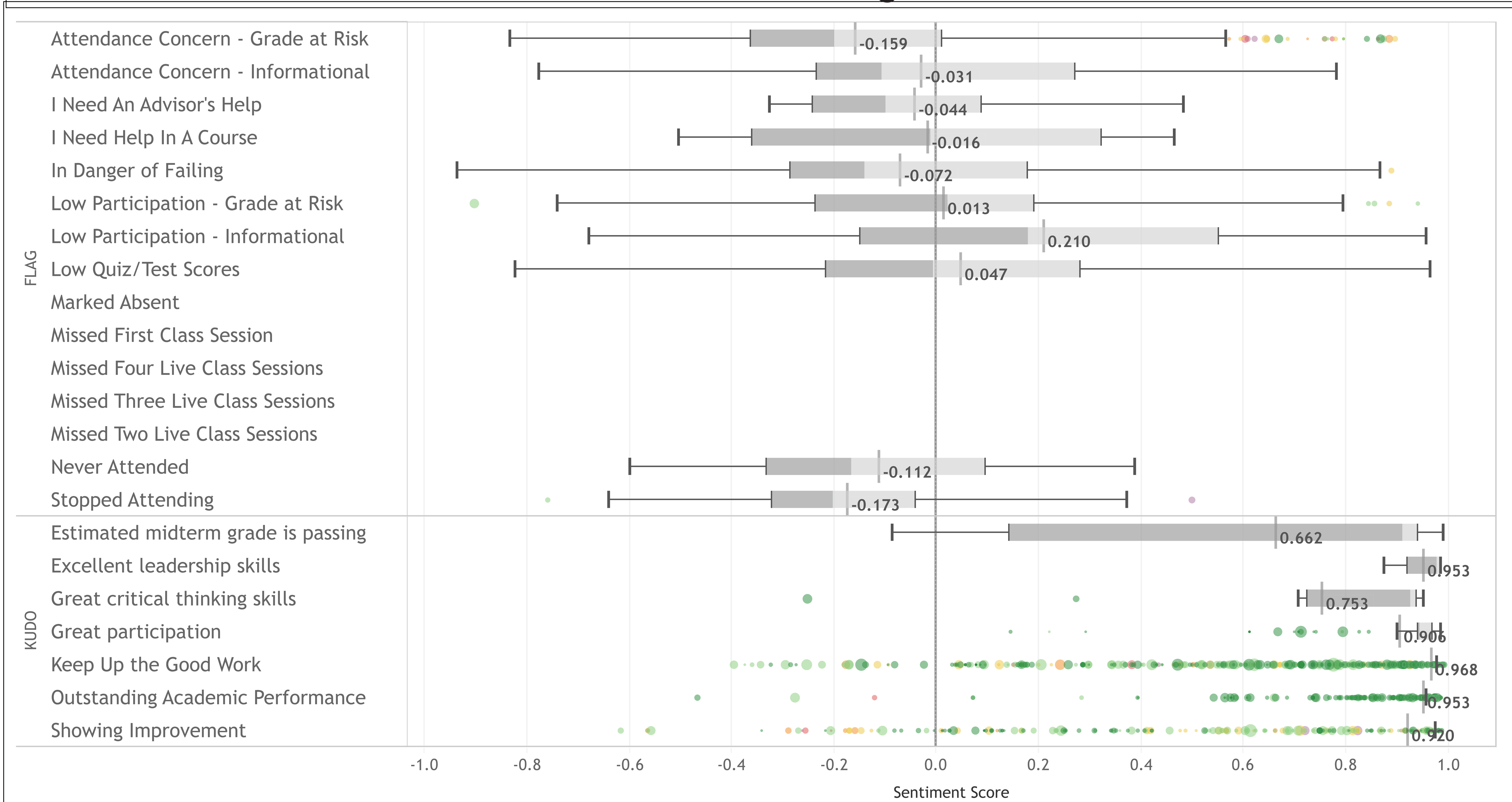
Scope and Tools
Timeline: late Fall 2023 to Spring 2024

Scope of data
- Early alert data (flags/kudos) [Fall '22 to Spring '23]
- Grades and enrollment data [Fall '22 to Spring '23]

Tools used:
- Python: data processing, data modeling
- R: data modeling
- Tableau: visualization, data exploration
- Models: distilbert-base-multilingual-cased-sentiments-student, association rule mining

New processes
- Utilize extensive note data from flags and kudos within a machine learning operation.
- Capture more diverse scenarios and patterns for more comprehensive insights.
- Train models effectively to yield higher predictive accuracy.
- Use down sampling techniques to balance class distribution.

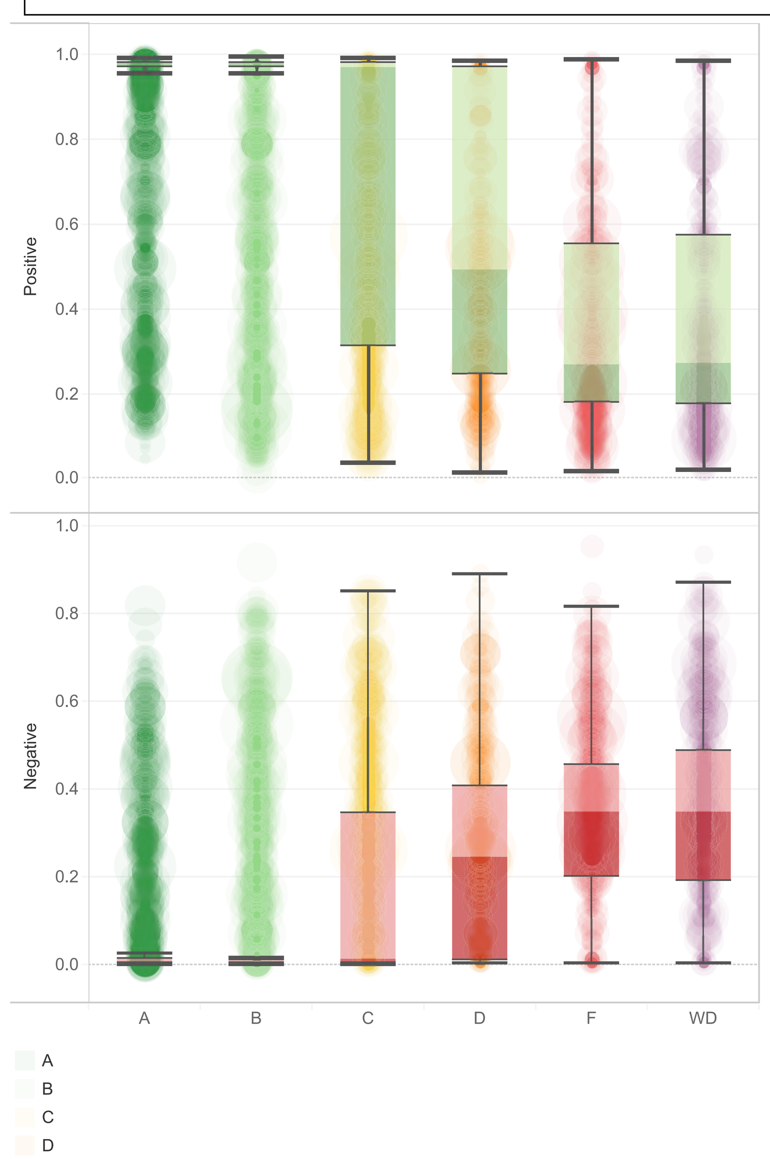
Sentiment data from flag/kudo comments



Key text analytic metrics

- **Sentiments**
  The overall tone conveyed in a text, speech, or piece of writing.
- **Emotions**
  The overall emotion expressed in a text, speech, or piece of writing
- **Lexical Diversity**
  Lexical diversity is a measure of how many different words appear in a text.
- **Lexical Density**
  The proportion of lexical words in a text compared to the total number of words.
- **Sentence specificity**
  The level of detail or precision contained within a sentence.

Sentiment visualization and grade outcomes



Model results

- **Random forest** exhibits better overall performance, making it the most suitable model for this dataset.
- **Decision tree** show similar performance levels.
- **Support vector machine learning (SVM)** slightly outperforming Random Forest and Decision Tree. However, as this data set contains a lot of noise, such as overlapping target classes, SVM will not perform as well when scaled.

		Cross Validation (3-Folds)			
	Class Imbalace	Accuracy	Precision (Weighted)	Recall (Weighted)	FI-Score (Weighted)
Decision Tree	None	0.59	0.67	0.59	0.56
	Up Sampleing	0.63	0.67	0.63	0.60
	Down Sampling	**0.63**	**0.66**	**0.63**	**0.60**
	SMOTE	0.60	0.68	0.60	0.57
Random Forest	Up Sampleing	0.63	0.66	0.63	0.61
	Down Sampling	**0.63**	**0.66**	**0.63**	**0.61**
SVM	Down Sampling	0.64	0.67	0.64	0.61
KNN	Up Sampleing	0.68	0.69	0.68	0.66
	Down Sampling	0.62	0.65	0.62	0.57

Kal Srinivas, Ph.D. ; Radell Roberts; ShawnMarie Parry; Samantha Trumble; Hopeton Smalling, Ed. S.

Taylor, L. (2012, July). Beyond Retention: Using Targeted Analytics to Improve Student Success. https://www.academicimpressions.com/how-early-alert-and-student-success-initiatives-fail/