2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
ASEE
Paper ID #37239

Predicting Academic Performance for Pre/Post-Intervention on Action-State Orientation Surveys

Prof. Ismail Uysal, University of South Florida

Dr. Ismail Uysal has a Ph.D. in Electrical and Computer Engineering from the University of Florida. He is an Associate Professor and the Undergraduate Director at the University of South Florida's Electrical Engineering Department. His research focuses on theory and applications of machine learning and machine intelligence for sensor applications.

Paul E. Spector
Dr. Chris S. Ferekides, University of South Florida
Mehmet Bugrahan Ayanoglu
Rania Elashmawy, University of South Florida

Received a B.S. degree in electronics and communication engineering from Arab Academy for Science, Technology and Maritime Transport, Cairo, Egypt, in 2012, and the M.S. degree in electrical engineering from the University of South Florida (USF), Tampa, Florida, USA, in 2019. She is currently pursuing the Ph.D. in electrical engineering with the University of South Florida, Tampa, Florida, USA. Her research interests include smart agriculture, precision agriculture, and time-series data.

Predicting Academic Performance for Pre/Post Intervention on Action-State Orientation Surveys

Ismail Uysal, Paul Spector*, Chris Ferekides, Mehmet Ayanoglu & Rania Elashmawy
Dept. of Electrical Engineering, *Dept. of Psychology, University of South Florida
Tampa, Florida, United States

Abstract

The objective of this study is to analyze responses to a survey that assesses behavior and cognitive processes linked to academic performance of freshman and junior students and explore the individual survey responses as potential predictors of the students' academic performance using statistical methods including machine learning algorithms and related data analytics.

The datasets used for this objective include undergraduate students at the University of South Florida registered as part of the following cohorts:

- Spring 2021 Cohort (1) – Electrical Engineering Juniors
- Spring 2021 Cohort (2) – General Engineering Freshman
- Spring 2021 Cohort (3) – Psychology Majors
- Fall 2021 Cohort (1) – General Engineering Freshman, and
- Fall 2021 Cohort (2) – Psychology Majors.
- Fall 2022 Cohort (1) – Electrical Engineering Juniors

In addition to the direct responses, we generated functions to represent features and attributes for each response, such as efficacy, habits, hesitation, preoccupancy, volatility, engagements in curricular and extracurricular activities. The student populations from all cohorts were combined to create a master survey list. Binary categories have been defined as academic failure (GPA < 2.0) or not (GPA > 2.0) based on the self-reported GPA by the students. Since students with GPA > 2.0 have constituted a much larger percentage of the population, we approached this problem as one-class anomaly detection, a well-defined area of machine learning. We implemented six different machine learning algorithms including K-Means clustering, deep neural networks (DNNs), principal component analysis (PCA), Guassian process regression (GPR), one-class autoencoders (OCAE) and one-class support vector machines (OCSVM) to identify if a student is academically successful (GPA > 2.0) or not. The highest accuracy topologies were OCAEs and OCSVMs.

The ML models were trained using only the students with GPA > 2.0 with randomly selected survey questions. Once a model has been created and trained, we tested the architecture using

survey responses that were never seen by the model. This test dataset consisted of a subsample of students with GPA > 2.0 and all the students with GPA < 2.0. As a reminder, up until this point the model had never seen any survey data from students with GPA < 2.0. The expectation was that the model would accurately categorize these test instances as anomaly samples based on the reconstruction error comparisons with the normal samples.

The train/test procedure was repeated for thousands of combinations of 18 randomly selected survey questions from the 60-question survey to find out which questions more consistently result in better predictions of academic failure. The best performing 18 feature groups were recorded for the top-10 most accurate classification scenarios using the area-under-curve (AUC) score as an indicator of percentage-based performance for binary classification tasks (i.e., is the student's GPA < 2.0 or not) specifically for heavily biased datasets such as this. For instance, a score of 0.744 means that approximately ~%74.4 of the time we can identify a student's likelihood of having a lower GPA using the survey questions used for that specific combination. After analyzing the performance results, and looking at the top performing combinations, we observed that the responses to questions such as 59, 26, etc. have disproportionally larger representations among the more accurate categorizations. Most of these questions involve study habits (as expected), but some also include extracurricular activities such as involvement in student clubs including IEEE as and on-campus housing activities.

Introduction

There are many factors that have been linked to academic success of college students. Although the importance of cognitive ability has been well established (Richardson et al., 2012), less clear is the potential impact of cognitive control processes (how people maintain effort toward goals) that impact behavior linked to academic performance. Our focus in this presentation is exploring how the cognitive control process of action-state orientation (Kuhl, 1992) of students would link to academic behavior that is important for academic success. Our focus here is the link of two domains of behavior, study habits and engagement in extracurricular activities, with grade point average (GPA). To that end we utilized a machine learning approach to identifying critical behaviors that link to college student GPA.

Action-state model

The theory of action-state orientation (Kuhl, 1992) explains how the achievement of goals depends upon the ability to self-regulate goal-relevant behavior. Action-state orientation itself reflects individual differences in how well people can regulate actions that are necessary to accomplish goals. People who are action-oriented engage cognitive control processes that enable them to maintain effort to progress toward meeting goals. An action-oriented student can set

academic goals, devise strategies to accomplish those goals, and execute those strategies. State-oriented students might set the same academic goals and devise the same strategies, but they struggle to maintain the cognitive control needed to turn plans into success. There are three ways in which the cognitive control of state-oriented individuals breaks down.

- Hesitation: Students have a hard time getting started. They procrastinate rather than engage with schoolwork.
- Preoccupation: Students can have a difficult time returning to a task after interruption.
- Volatility: Students can have a difficult time staying focused on a task; they get bored and find a more interesting activity rather than schoolwork.

There is not a lot of research on the behavioral strategies that people in general, or students in particular might use to overcome state orientation. We theorize that hesitation and volatility can best be addressed by setting short-term goals. For example, a student who struggles to get started reading a chapter might be more successful if the short-term goal is to read a limited number of pages. A student who has a hard time staying focused while bored would also do better with a goal to read a limited number of pages rather than an entire chapter. Preoccupation can be addressed by eliminating distractions, such as shutting off cell phones while studying.

Behaviors Relevant to Academic Success

There are two classes of behavior that have been linked to academic success. Extracurricular engagement is participation in activities outside of the classroom. It has been linked to a variety of academic success indicators including GPA (Bakoban & Aljarallah, 2015), graduation (Flynn, 2014), and post-graduation earnings (Hu & Wolniak, 2013). Study habits are the strategies that students use to accomplish their coursework. It consists of behaviors such as finding a quiet place to study and avoiding all-nighters. Studies have linked study habits to academic success (Nonis & Hudson, 2010).

A limitation to most studies of engagement and study habits is that they use measures that combine items into dimension scores. Because these measures include items of different behaviors that are not interchangeable, they are best considered formative scales (Edwards & Bagozzi, 2000). Although there is value in relating overall dimension scores to important outcomes like GPA, doing so makes it difficult to offer precise advice to students about which behaviors to adopt to make the most efficient use of their efforts. Thus, this study investigated individual behaviors by analyzing results at the item level.

Data Processing

Survey data as opposed to standardized scientific data presents its own challenges. Scientific datasets are more likely to include a higher level of correlation between the measured and predicted features. Survey data on the other hand is more likely to have outliers and anomalies, making the algorithm fitting or training a difficult task. After using some of the more traditional approaches including standard machine learning methods such as deep neural networks, we came to realize that a more experimental approach in preconditioning and filtering of our feature selection process needs to be implemented.

The preparation of the dataset consists of first cleaning the anomaly inputs such as non-numerical values entered in numerical fields, or out of range values such as GPAs below 0 or above 4. About 60 features are used in the study, which correspond to the 60 questions asked in the survey. In addition to these features, we have artificial responses generated from functions that use the responses to specific questions, such as efficacy, habits, hesitation, preoccupancy, volatility, and engagements in curricular and extracurricular activities. Efficacy feature uses the responses to the questions 1 through 7, while "habits" feature uses 8 through 29, "hesitation" uses 38 through 45, "preoccupancy" uses 30 through 37, "volatility" uses 46 through 50, "engagements in curricular" uses 51 through 54, and lastly "engagements in extracurricular" uses the responses to the questions 55 through 59.

The dataset responses are then aligned where different cohorts are concatenated, saved, and then normalized using MinMax scaler. The scaling is done to prevent biasing in supervised learning models which can occur by having a feature, larger in magnitude, impact the training more than the others because of not being in the same numerical range as other features. Minmax scaler simply brings each feature to the same range while maintaining the ratio between the instances of the dataset for that specific feature.

To get an idea on the similarity of the features of our dataset we used Pearson's Correlation Coefficient. Pearson's method attempts to fit a plot function to represent the similarity between any given two arrays, which are features in our case. The value of this coefficient ranges from 0 to 1, where highly related features have a value closer to 1 and as the data gets dissimilar, values drop down towards 0. As shown in Figure 1, it is safe to say that there is minimal similarity between the features of our dataset where the diagonal – which indicates self-correlation between features which is equal to 1 – dominates the color palette. Another interesting observation is the similarities between the artificial features (i.e., habits, hesitation, preoccupation, etc.) and direct responses are represented as they are calculated directly from these features.

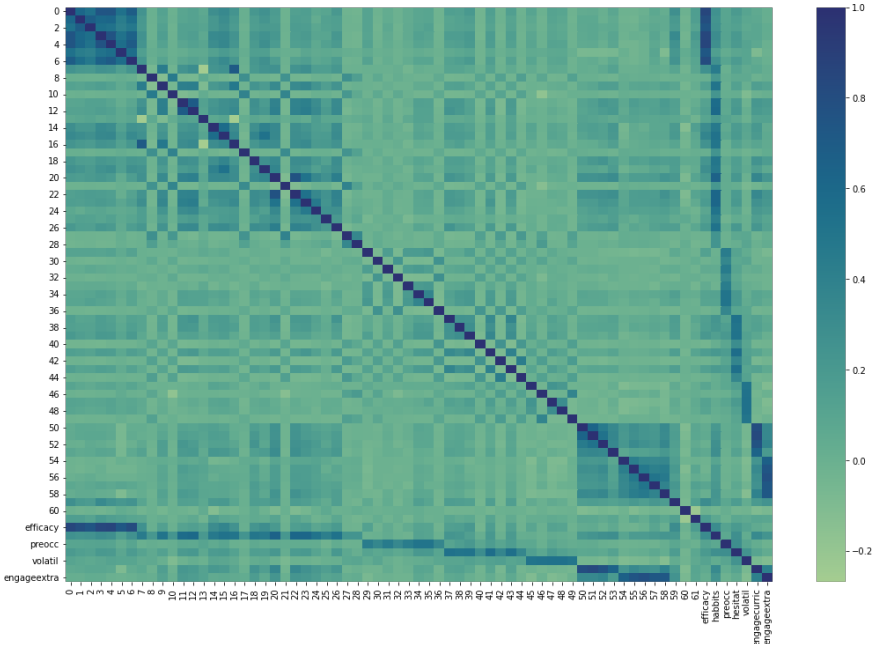

Figure 1: Pearson's Correlation Coefficient map including all survey questions in addition to artificial features

Objectives

The fundamental objective of the study was to:

i) first identify if action-state survey questions can be used as predictors of academic failure and later,

ii) to categorize the importance of these questions when it comes to academic failure.

To achieve these objectives, we have defined the following simple binary categorization based solely on self-reported GPA.

GPA Value	Category Representation
GPA > 2.0	Normal Sample
GPA ≤ 2.0	Anomaly Sample

Table 1: Categorization of Dataset

The reason why GPA values less than or equal to 2.0 were classified as anomaly samples is because of the proportion of these responses compared to the rest of the surveyed popular (i.e., less than 5%) In fact, figure 2 demonstrates the distributions of self-reported GPAs around 2.0, 3.0 and 4.0 values which clearly indicates the vast majority of the samples are 2.5 or higher.

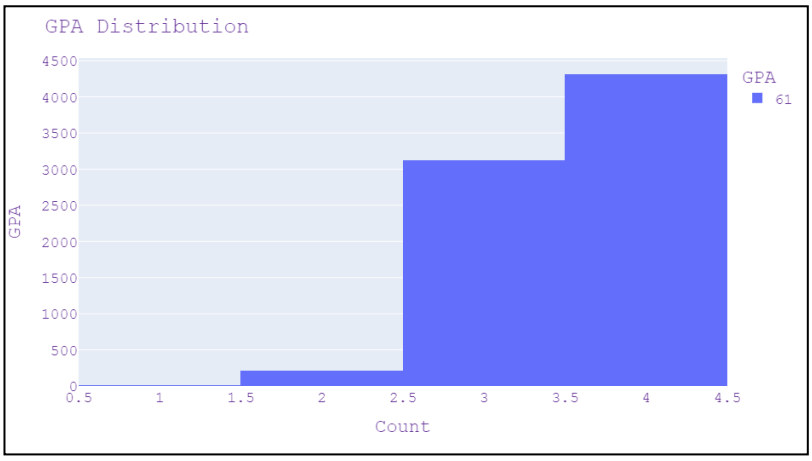

Figure 2: Self-reported GPA histograms

When we tried different thresholds as boundaries for "anomalies", the detection performances of both algorithms (detailed in the methodology section) have yielded unacceptable trade-offs for true versus false positive rates as shown in figure 3 below. The blue curve has the only distinct difference from the 0.5 baseline compared to other thresholds justifying our selection of 2.0 as the anomaly boundary.

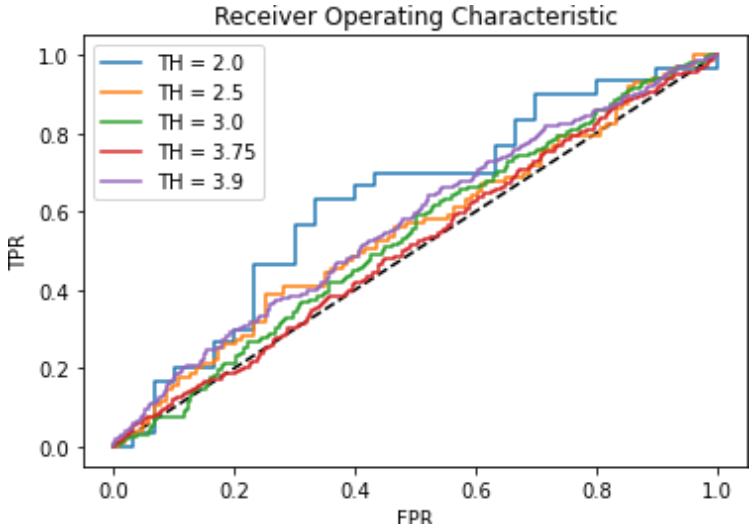

Figure 3: True-positive-rate versus false-positive-rate (Receiver-Operating-Characteristics curve) for the autoencoder model for different GPA thresholds.

Methodology

Two fundamentally different machine learning algorithms have been used in the study to capture both conventional and modern approaches to predictive analytics. Ultimately, the learning algorithm was formulated an outlier detection method to find the anomalies in the dataset using

machine learning methods including one-class autoencoders and one-class support vector machines.

The one-class autoencoder works by building a latent space representation of survey responses labeled as “normal” where the reconstruction error between the input and output is minimized. The hypothesis is that when “anomaly” samples are presented, the reconstruction error would be higher which would then signal the existence of an anomaly. The structure of a basic one-class autoencoder is shown in figure 4 below.

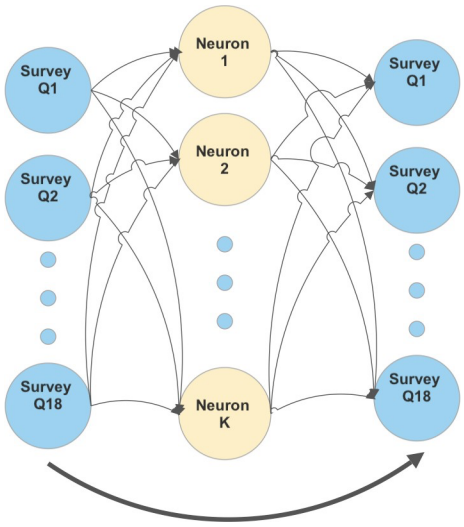

Figure 4: One-class autoencoder with K-dimensional latent space using 18 select questions from the survey for detection.

An interesting fact from figure 4 is that instead of the entirety of 60 questions, only 18 questions were selected to be used in training. The justification is that in this current study limited conditioning has been performed on the input features utilized in our ML methods. Therefore, we opted to investigate different selections of input features and identify those that yield optimal performance in the autoencoder outlier categorization approach. We randomly chose 18 features based on the number of features determined from our principal component analysis with a 5% threshold, as the baseline for comparison. The autoencoder model was then executed using these selected features, and the resulting area-under-the-curve (calculated from score and corresponding features (i.e., question numbers) were stored in a data frame for 20,000 randomizations.

In addition to the autoencoder model, using similar feature randomization, we also implemented One Class Support Vector Machine (OCSVM) as another very effective method used in anomaly detection. OCSVM is an unsupervised method, meaning the model does not know the labels of the training data, which in this case are “normal” vs. “anomaly”, but still performs a clustering,

boundary creation algorithm to find the anomalies in the dataset. It not only predicts the output categorization of the input data, but also assigns a confidence factor by calculating the distance of the instance to the decision boundary on the data plane. This method provides significantly faster training compared to the autoencoder approach.

Results

For most anomaly detection exercises, area-under-the-curve, or AUC score is used as an indicator of performance rather than simple accuracy. Receiver operating characteristic (ROC) curve is used to calculate the AUC score. The ROC curve is a plot of true positive rate (TPR) against false positive rate (FPR) at various threshold settings. The TPR is the proportion of positive samples that are correctly classified as positive, while the FPR is the proportion of negative samples that are incorrectly classified as positive. The AUC score is equal to the area under the ROC curve, which ranges from 0 to 1 where 1 indicates the perfect score (100% accuracy with no false negatives or positives) and 0.5 indicates the worse performance (basically a coin-flip when it comes to predicting an anomaly.

Tables 2 and 3 display the highest AUC scores achieved along with the associated features used for both autoencoder and OCSVM approaches among 20,000 different combinations which has been trained/validated and tested across the entire dataset.

	Combinations	AUCScores
319	[4, 27, 47, 16, 26, 46, 65, 3, 53, 40, 61, 36, 45, 11, 57, 29, 34, 59]	0.744444
157	[14, 65, 10, 46, 61, 15, 21, 1, 7, 67, 12, 59, 28, 47, 52, 56, 20, 18]	0.741111
441	[17, 35, 42, 53, 8, 49, 26, 14, 3, 31, 67, 63, 4, 39, 47, 9, 61, 37]	0.704444
144	[2, 62, 24, 56, 38, 14, 15, 29, 10, 40, 32, 8, 44, 46, 60, 0, 43, 36]	0.701111
225	[35, 38, 64, 18, 37, 10, 27, 28, 20, 22, 25, 46, 26, 61, 33, 59, 43, 62]	0.684444
71	[41, 13, 49, 11, 39, 20, 4, 1, 19, 28, 8, 59, 62, 34, 30, 3, 17, 56]	0.675556
496	[52, 28, 38, 10, 7, 37, 60, 26, 6, 27, 49, 2, 53, 14, 47, 8, 64, 4]	0.673333
471	[59, 26, 0, 28, 33, 7, 25, 46, 61, 54, 9, 23, 34, 67, 22, 11, 40, 56]	0.672222
456	[67, 43, 3, 40, 12, 62, 51, 29, 18, 15, 9, 59, 28, 66, 25, 20, 0, 46]	0.667778
69	[63, 52, 9, 48, 37, 5, 43, 62, 0, 13, 55, 31, 16, 53, 20, 25, 56, 66]	0.667222
485	[29, 30, 27, 3, 8, 33, 61, 5, 10, 63, 32, 37, 7, 26, 67, 36, 22, 28]	0.664444

Table 2: Feature combinations and AUC scores for the autoencoder model

	Combinations	AUCScores
6698	[58, 19, 29, 55, 39, 53, 59, 63, 15, 62, 47, 22, 41, 54, 4, 35, 60, 38]	0.728552
15284	[54, 51, 41, 58, 46, 13, 53, 60, 29, 42, 6, 55, 50, 23, 30, 59, 21, 35]	0.720157
2175	[38, 23, 67, 14, 50, 56, 29, 19, 53, 54, 42, 36, 46, 34, 62, 59, 16, 9]	0.709118
2641	[10, 54, 60, 42, 58, 66, 46, 0, 15, 7, 18, 41, 47, 31, 37, 22, 53, 26]	0.70873
16639	[22, 30, 67, 29, 58, 41, 19, 34, 50, 40, 43, 36, 54, 42, 63, 53, 27, 59]	0.708366
12607	[66, 37, 40, 61, 55, 58, 26, 65, 7, 53, 29, 46, 39, 44, 24, 45, 36, 31]	0.706473
272	[41, 67, 29, 7, 19, 21, 59, 38, 50, 36, 63, 46, 18, 15, 45, 56, 14, 42]	0.706012
5442	[20, 67, 46, 13, 6, 26, 65, 38, 64, 18, 16, 39, 31, 22, 15, 29, 53, 36]	0.701985
16583	[36, 22, 54, 9, 59, 19, 66, 20, 2, 46, 29, 6, 39, 17, 57, 55, 40, 30]	0.69934
3357	[53, 67, 15, 39, 60, 29, 9, 43, 38, 14, 46, 47, 51, 26, 19, 41, 1, 36]	0.698054
5302	[21, 36, 46, 34, 35, 42, 63, 41, 45, 32, 6, 60, 40, 53, 31, 47, 38, 22]	0.696332
3235	[29, 51, 54, 25, 67, 15, 26, 44, 40, 60, 14, 20, 22, 37, 31, 53, 42, 10]	0.694997
14458	[59, 54, 30, 36, 16, 47, 31, 67, 66, 7, 5, 58, 17, 32, 10, 4, 29, 19]	0.694196

Table 3: Feature combinations and AUC scores for the OCSVM model

Area-under-curve scores indicate the performance for binary anomaly classification (is GPA < 2.0 or not) where the top score 0.744 means ~%74.4 of the time we can identify a student's likelihood of falling under the threshold using the survey questions listed under that combination. Moreover, when looked carefully, similar feature combinations appeared in both autoencoder and OCSVM methods' best performing results. Figure 5 below shows how often each feature (i.e., survey question) appeared in the top-20 results for both algorithms.

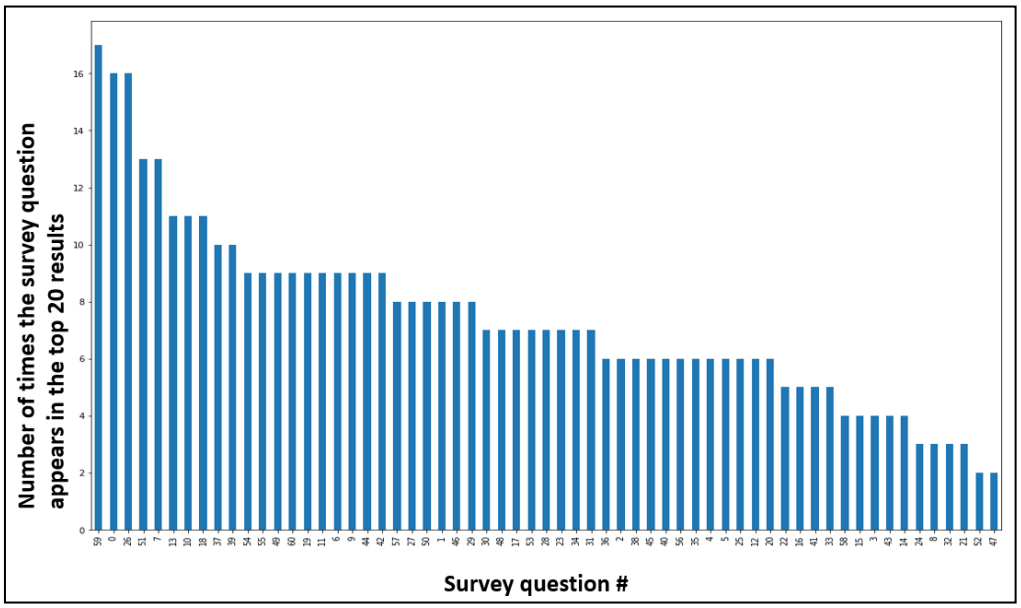

Figure 5: Most frequency survey questions to provide the best AUC scores across algorithms.

Applying the randomly selected 18 features approach to our model that uses the OCSVM resulted in performance results similar to the autoencoder approach. The advantage is that this methodology is less time and memory intensive compared to the autoencoders.

Conclusions and Future Work

Regardless of the algorithm used, on average there is approximately a 3 in 4 chance (~75%) to predict if a student is academically in danger of failure based on the responses submitted to action-state surveys. More importantly, some survey questions, specifically 59, 26, 51 and 7 have disproportionally larger representations among the more accurate categorizations. Most of these questions involve study habits such as allowing friends to disrupt studying or doing all-night study sessions for preparation but some of them also include extracurricular activities such as involvement in student clubs including IEEE as well as on-campus housing activities. A closer look is necessary to find out the specific contributions for each of these questions and that there is no algorithm training bias – although the latter is unlikely as two structurally very different algorithms were used in training.

For future work – we will focus on the impacts of student interventions in creating quantifiable differences in their survey responses by asking the following question: can a model trained on pre-intervention action state surveys be used to identify the level of improvement in a student's state of mind post intervention using objective changes in prediction performance. We have started the preliminary work on this where the first step involves identifying students across their survey responses since the surveys are anonymous. To ensure accurate data matching, the team employed identity survey questions to establish a clear connection between students who took the survey before and after. This involved using key pieces of information, such as gender, ethnicity, month of birth, city of birth, middle name initial, and high school attended, from the demographic section. These questions were chosen carefully to provide a comprehensive picture of everyone in the dataset and prevent errors or discrepancies while keeping the survey anonymous.

To match the high school names and cities of birth automatically, the team used the Python library FuzzyWuzzy. This library matches strings and calculates the differences between words or phrases. The team utilized two modules from the library: fuzz.partial_ratio and fuzz.token_sort_ratio. Fuzz.partial_ratio calculated the similarity score for abbreviated or shortened forms of the high school name or city of birth, such as "NY High School" and "New York High School". Fuzz.token_sort_ratio was used for instances where the order of the words

in the name differed, such as "New York High School" and "High School New York". By using these identity survey questions, the team effectively matched the data, enabling more in-depth and accurate analysis of the collected information. The flowchart in figure 6 below demonstrates the matching algorithm which we have shown to work with greater than 95% accuracy (when compared to a human manually matching survey responses).

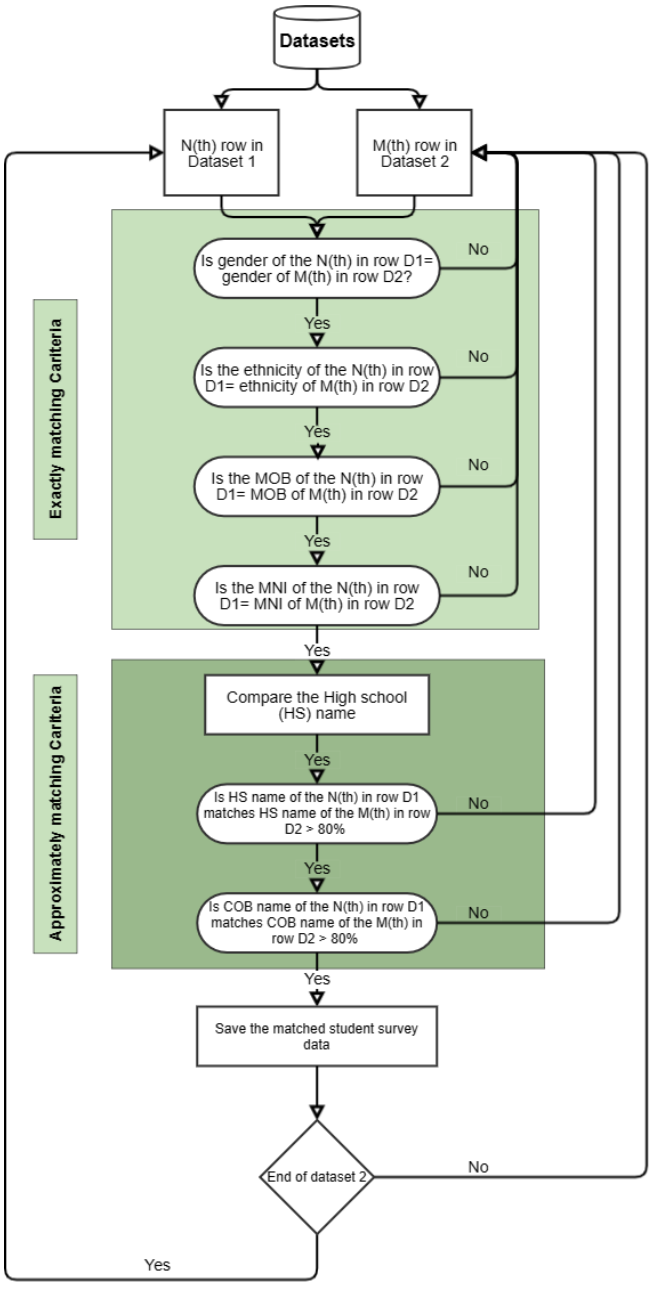

Figure 6: Logical flow-chart used in matching student survey responses across different semesters/cohorts for pre-post intervention.

Our initial results are promising. We matched approximately 100 students between pre and post intervention surveys (i.e., same student taking the survey before and after intervention) and used both pre and post-intervention survey responses in predicting their academic success (or failure). The average AUC scores of pre-intervention surveys when used on a model trained solely with pre-intervention data was 0.81 whereas the AUC scores of post-intervention surveys when used on the same model was 0.51. In other words, the model can successfully identify a student who is likely to fail academically before the intervention but identifies the same student as academically successful (even though their GPA has not changed) based on their survey responses after intervention. This indicates (but does not yet prove) that there is measurable and quantifiable difference in how the same students respond to the survey before and after intervention which should, hopefully, ultimately lead to better outcomes.

References

Bakoban, R., & Aljarallah, S. (2015). Extracurricular Activities and Their Effect on the Student's Grade Point Average: Statistical Study. *Educational Research and Reviews*, *10*(20), 2737-2744.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155-174.

Flynn, D. (2014). Baccalaureate attainment of college students at 4-year institutions as a function of student engagement behaviors: Social and academic student engagement behaviors matter. *Research in Higher Education*, *55*(5), 467-493. https://doi.org/10.1007/s11162-013-9321-8

Hu, S., & Wolniak, G. C. (2013). College student engagement and early career earnings: Differences by gender, race/ethnicity, and academic preparation. *Review of Higher Education: Journal of the Association for the Study of Higher Education*, *36*(2), 211-233. https://doi.org/10.1353/rhe.2013.0002

Kuhl, J. (1992). A theory of self-regulation: Action versus state orientation, self-discrimination, and some applications. *Applied Psychology: An International Review*, *41*(2), 97-129. https://doi.org/10.1111/j.1464-0597.1992.tb00688.x

Nonis, S. A., & Hudson, G. I. (2010). Performance of college students: Impact of study time and study habits. *Journal of Education for Business*, *85*(4), 229-238. https://doi.org/10.1080/08832320903449550

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol Bull*, *138*(2), 353-387. https://doi.org/10.1037/a0026838