The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #43785

Tracking and Predicting Student Performance Across Different Semesters with Matched Action-State Orientation Surveys and Interventions

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 in education, healthcare and smart agriculture including machine intelligence for sensor applications.

Mehmet Bugrahan Ayanoglu, University of South Florida
Dr. Rania Sherif Elashmawy, University of South Florida

Dr. Rania Elashmawy has a Ph. D in Electrical Engineering from the University of South Florida, USA. Her research interests include smart agriculture, precision agriculture, and time-series data.

Rifatul Islam, University of South Florida
Paul E. Spector, University of South Florida
Dr. Chris S. Ferekides, University of South Florida

Tracking and predicting student performance across different semesters with matched action-state orientation surveys and interventions

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

Abstract

This paper presents the second year results of the work supported by the National Science Foundation's Revolutionizing Engineering Departments (IUSE/PFE: RED) Program under the project titled "IUSE/PFE:RED: Breaking Boundaries: An Organized Revolution for the Professional Formation of Electrical Engineers." Specifically, this part of the study looks at action-state orientation and its impacts on student success. The first-year results were presented at the 2023 ASEE Conference in Baltimore, MD with the academic paper titled *"Predicting Academic Performance for Pre/Post-Intervention on Action-State Orientation Surveys"* for further reference (Uysal, 2023). The objective of the first phase of the study was to find out how survey responses could be used to predict whether a student could be considered at-risk for failing academically. The objective of the second phase discussed in this article is to analyze and quantify ***the effects of <u>in-class interventions</u> on student study habits*** and, ultimately, their academic performance using action-state orientation surveys as engineering students progress in their respective curriculum.

While these surveys are anonymous, it is crucial to be able to track changes in academic performance for individual students across multiple semesters as they go through the various stages of their academic program (in this case, Electrical Engineering). As part of the second phase, we developed a powerful matching method that can automate the demographic information matching in the background with Python libraries to ensure sustainable analysis as the data collected from both new and ongoing students naturally grew larger over the past several years. Ultimately, we were able to match a total of 840 unique students based on their self-provided information such as gender, month of birth, ethnicity, and high school names across 2148 unique survey responses collected in 5 different academic semesters: Spring 2021, Fall 2021, Spring 2022, Fall 2022, and Spring 2023. Beyond the scale of data, which is unprecedented for this kind of survey, we were able to significantly boost the prediction performance of our machine learning algorithms from 74.4% reported in the previous study for a simpler question (i.e., is this student's GPA less than 2.0? - a more apparent anomaly) to

This material is based upon work supported by the National Science Foundation under Grant No. 2022299.

82.6% for a more challenging question (i.e., is this student's GPA higher or lower than 3.33? - a more subtle distinction).

To accomplish this, we leveraged sophisticated machine learning classifiers before settling on the random forest classifier with feature elimination, thanks to the increasing size of the collected data from the newly added surveys. The students in the dataset were split into two groups based on their GPA such that the method can learn from survey responses to correctly identify the category (high or low GPA) where k-fold (10) cross-validation was used to ensure robust and repeatable accuracy metrics were obtained. The dataset was further split into two partitions by classifying survey responses as pre-intervention and post-intervention, where 921 unique responses were classified as post-intervention (*POST*), and 1227 unique responses were classified as pre-intervention (*PRE*). Using this information, a new predictor was trained using only the *PRE* dataset and was tested on both the *PRE* and the *POST* datasets. The hypothesis was that the new empirical model would perform worse with more false-positives on the *POST* dataset due to newly acquired and hopefully improved study habits after the interventions. Our results show a 35% increase in prediction error when the same algorithm is tested on the POST student population and more importantly a corresponding 24% increase in the false-positive rate which indicates that the interventions are working at the population level where students adopt study habits that outperform their current academic performance as likely indicators.

Introduction

Numerous factors contribute to the academic success of college students. While the significance of cognitive abilities has been clearly recognized (Richardson et al., 2012), the role of cognitive control processes, i.e., how individuals persist in their efforts towards their academic goals, and their influence on academically relevant behaviors remains less understood. In this study we concentrate on examining the relationship between the students' cognitive control, specifically their action-state orientation (Kuhl, 1992), and behaviors that are crucial for academic success. Specifically, we investigate how study habits and participation in extracurricular activities correlate with students' grade point average (GPA). To achieve this, we employed sophisticated data parsing and machine learning tools to identify the critical behavioral links to college student GPA.

Action-state model

The concept of action-state orientation, as initially proposed by (Kuhl, 1922), discusses how achieving objectives is closely linked to self-regulating behavior pertinent to those objectives.

This orientation highlights how each individual has varying levels of abilities to manage actions required for goal attainment. Action-oriented individuals are more adept at deploying cognitive control processes to sustain the effort needed for goal progression. For instance, an action-oriented individual can properly establish academic objectives, plan methods for achieving these goals, and implement these methods effectively to achieve said objectives. Conversely, state-oriented individuals may identify similar academic objectives and formulate similar plans but face challenges in sustaining the necessary cognitive control to turn these plans into completed achievements. There are three common ways in which the cognitive control of state-oriented individuals breaks down:

1. Hesitation: Students have a hard time getting started. They procrastinate rather than engage with schoolwork.
2. Preoccupation: Students can have a difficult time returning to a task after interruption.
3. 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 limited research on the behavioral tactics that individuals, especially students, can employ to overcome state orientation. We propose that short-term goal setting is an effective strategy for managing hesitation and inconsistency. For instance, a student facing difficulties in beginning to read a chapter could find it easier to start by reading just a few pages at a time. Preoccupation, on the other hand, can be addressed by minimizing distractions, like turning off cell phones during study sessions.

Behaviors Relevant to Academic Success

Two types of behavior have been identified as key to academic success. The first, extracurricular engagement, involves participating in activities beyond just the classroom setting. This engagement has been associated with various indicators of academic success, including GPA (Bakoban & Aljarallah, 2015), degree completion (Flynn, 2014), and even future earnings (Hu & Wolniak, 2013). The second, study habits, refer to the methods students employ to manage their academic work, including practices like seeking a quiet study environment and avoiding cramming sessions. Research has shown a link between effective study habits and academic achievement (Nonis & Hudson, 2010).

There exists a common limitation in studies on both extracurricular engagement and study habits through reliance on composite measures that aggregate diverse behaviors into single scores. Given that these aggregated measures mix various behaviors that are not directly comparable, they are more accurately described as formative scales (Edwards & Bagozzi, 2000). While associating overall scores with key numerical outcomes like GPA is useful, it may hinder the provision of specific guidance to students on which behaviors most efficiently contribute to their success. Consequently, our study initially focused on examining individual behaviors through an item-level analysis.

Data Collection and Preparation

Surveys were conducted over approximately two years across 5 academic semesters: Spring 2021, Fall 2021, Spring 2022, Fall 2022, and Spring 2023, and a range of different courses and student cohorts which generally took the survey multiple times both prior to and after the so-called *action-state interventions* in the classroom to improve study habits. A representative figure for some of the survey questions in measuring the action-state orientation of students is provided in Figure 1 below. In the end, we collected a total of 2148 survey responses.

The in-class intervention provided tips to students for overcoming state-orientation tendencies. Beginning early in the semester, the professor provided an overview of action-state orientation and discussed tips for improving performance in classes and beyond. Tips included setting mini goals for daily accomplishments, avoiding distractions (e.g., turning off cell phones), taking periodic breaks, and spreading exam preparation over time rather than cramming the night before.

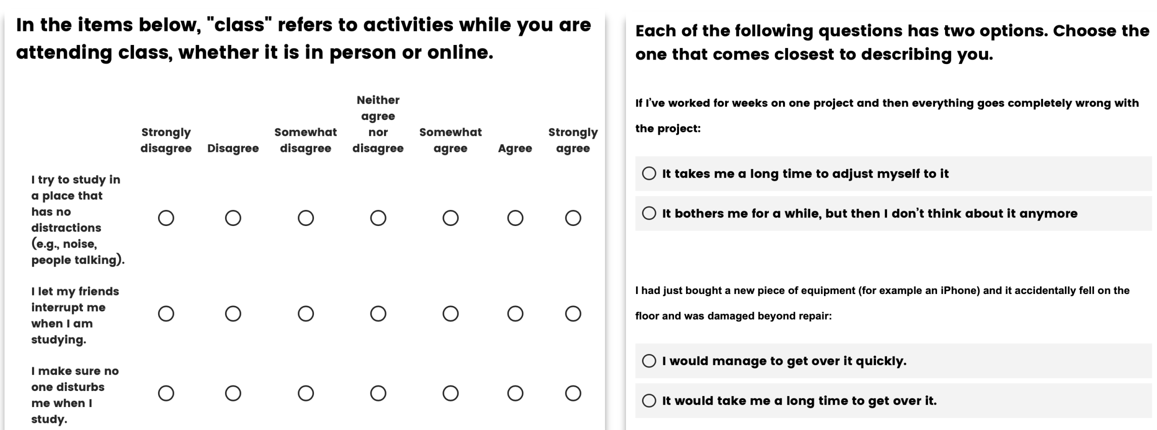


Figure 1: Surveys asked both scalar and binary questions on a range of topics including study habits both in and outside the classroom as well as extracurricular activities.

As detailed in (Uysal, 2023), the preparation of the dataset consisted 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 questions were common to all the surveys which were subsequently used as features in the machine learning models of the study. 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 responses in the dataset are first organized by aligning and concatenating different cohorts, after which they are saved and subsequently normalized using a MinMax scaler. This normalization process is crucial for mitigating potential biases in supervised learning models. Such biases can arise when features of larger magnitude disproportionately influence the model's training due to their numerical range not aligning with that of other features. The MinMax scaler addresses this issue by adjusting all features to a uniform range, thereby preserving the ratios among the dataset's instances for each specific feature.

Unlike the previous study, the output categorization has changed where the "academic success" is now defined as having a GPA of greater than 3.33 instead of 2.00 as shown in Table 1 below. Any sample above 3.33 was labeled as TRUE whereas any sample below 3.33 was labeled as FALSE for the purposes of classification labels. The main reason for changing the GPA threshold to 3.33 is the fact that it represents the median GPA point for the dataset and creates a balanced representation of both classes for the training/testing processes.

GPA Value	Category Representation
GPA > 3.33	TRUE
GPA ≤ 3.33	FALSE

Table 1: Categorization of Dataset

The main objective of the second phase of the study was to analyze the impacts of student interventions in creating quantifiable differences in their survey responses by answering the following questions:

1. Can a predictive model be trained on the survey responses with sufficient accuracy compared to the baseline (in this case 50% for a binary classification) in classifying student GPA groups as TRUE or FALSE?

2. If the answer is yes to question 1, does the model trained only on pre-intervention action

state surveys have quantifiable levels of difference in accuracy when tested on the post-intervention responses?

3. If the answer is yes to question 2, is this difference in accuracy reflected in explainable and modest changes in false-positive ratios between the models trained and tested on different populations?

To start answering these questions, we need to be able to identify and match students and their survey responses across different courses and semesters. This is not a straightforward task due to the fact the surveys are anonymous and there is no unique identifier to allow matching in the background. 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.

Since some of this demographic information is typed in, there are differences in responses to some of the questions, such as the high school names and cities of birth even by the same student which makes it challenging to conduct a trivial string search. Hence, the team used a Python library specifically designed for this purpose called *FuzzyWuzzy*. This library uses fuzzy logic to match strings and calculates a numerical difference between words or phrases using the modules fuzz.partial_ratio and fuzz.token_sort_ratio. Fuzz.partial_ratio calculates the similarity score for abbreviated or shortened forms of the high school's name or city of birth, such as "NY High School" and "New York High School" and Fuzz.token_sort_ratio was used for instances where the order of the word were different, such as "New York High School" and "High School of 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 below demonstrates the matching algorithm in detail which was shown to work with greater than 95% accuracy when compared to a trained expert manually matching survey responses in a smaller subset of the survey data.

Ultimately, we were able to match a total of 840 unique students based on their self-provided information such as gender, month of birth, ethnicity, and high school names across 2148 unique survey responses. These students were labeled with the naming convention of AAA (first unique individual), AAB, AAC, etc. and tagged with when they took that particular survey.

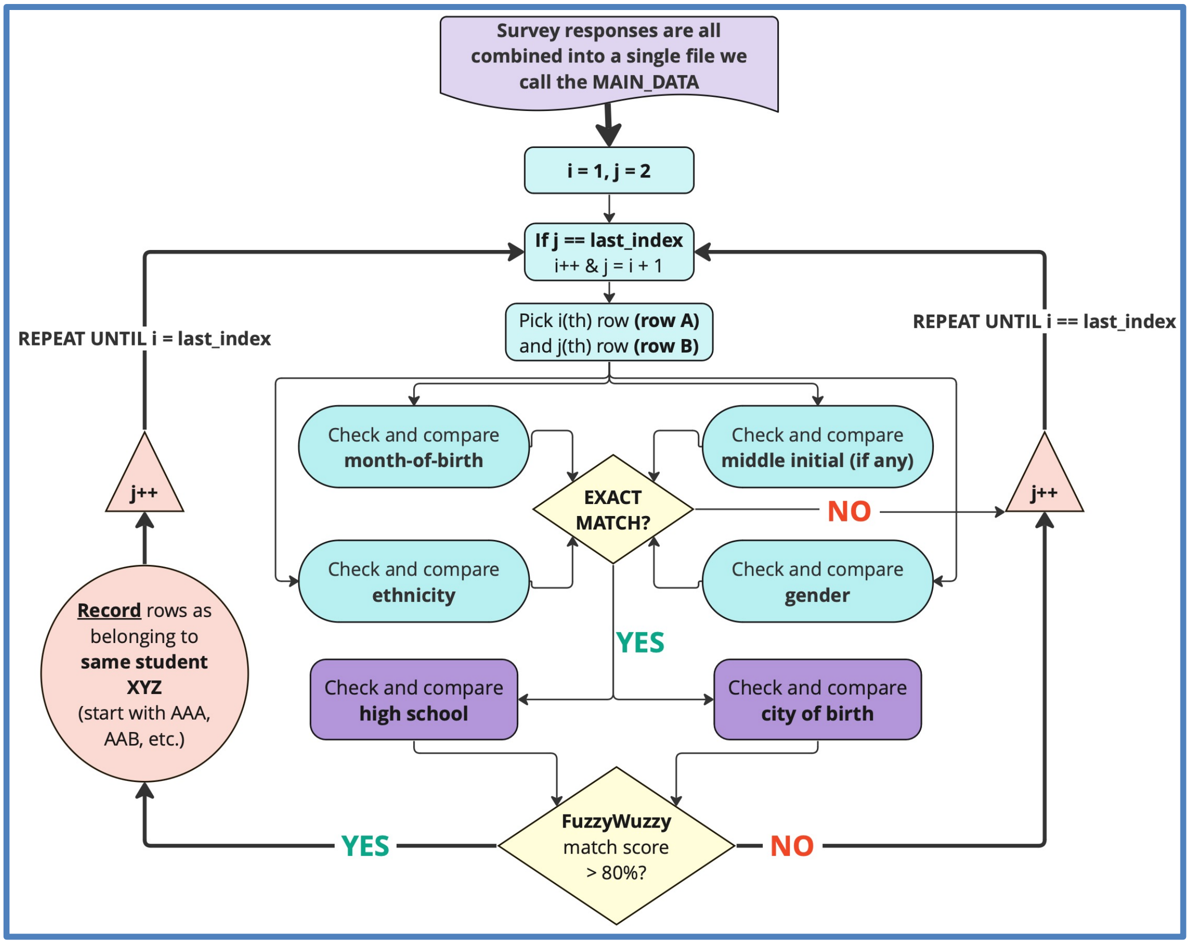


Figure 2: Logical flow-chart used in matching student survey responses across different semesters/cohorts for pre-post intervention.

We then proceeded to split the main dataset into two partitions labeled PRE and POST. To create the largest hypothesized difference between survey responses, we took only the **earliest** survey response recorded **before** any action-state intervention and only the **latest** survey response recorded **after** any action-state intervention. For instance, if the student AAA took two surveys in different semesters but both were taken after interventions, we excluded that student in this comparative study. Similarly, if the student AAB took multiple surveys across multiple semesters, such as Fall 2021, Spring 2022, Fall 2022, we only took the earliest PRE response (i.e., Fall 2021) and the latest POST response (i.e., Fall 2022). Tables 2 and 3 present a brief snapshot of how these newly partitioned datasets look like for both PRE and POST splits for the first six students and five questions. There are 549 unique students in each partition along with 60 questions for each survey response.

Identity	SurveyData	Q1	Q2	Q3	Q4	Q5
AAB	PRE_FALL21	6	5	5	6	6
AAC	PRE_FALL22	6	7	7	7	7
AAD	PRE_FALL21	5	6	6	5	5
AAF	PRE_FALL21	6	6	6	6	6
AAG	PRE_FALL21	6	5	5	6	6

Table 2: A sample collection for pre-intervention responses

Identity	SurveyData	Q1	Q2	Q3	Q4	Q5
AAB	POST_FALL22	6	7	6	6	6
AAC	POST_FALL22	7	7	6	7	7
AAD	POST_SPRING22	7	6	6	7	6
AAF	POST_SPRING22	7	7	7	7	7
AAG	POST_FALL22	6	7	6	6	6

Table 3: A sample collection for post-intervention responses

Methodology

In the previous study (Uysal, 2023) we observed that not all questions were equally affective in predicting the academic success of a student. Using autoencoders and support vector machines we were able to predict (~74.4% chance) if a student is academically in danger of failure based on the responses submitted to action-state surveys. In the previous study the academic success was defined as having a GPA of 2.0 or above as other GPA thresholds resulted in poor predictive performance.

In this study – we use a novel feature selection method developed specifically for this project as shown in Figure 3 below. In probability weighted feature selection (PWFS), a subset of features (in this case question responses) is selected randomly and the model is subsequently trained and validated. Based on the accuracy of the model on the validation set, each randomly selected feature is assigned a probability weight which makes it more or less likely for that feature to be selected on the next round of randomization. This allows for an evolutionary approach where not only the performance of the algorithm is empirically boosted but also the importance of features in predicting the desired outcome is quantified. The details of the feature selection algorithm are beyond the scope of this text and the reader is encouraged to contact the authors for a sample script and more information on how to apply this algorithm to other survey-based data.

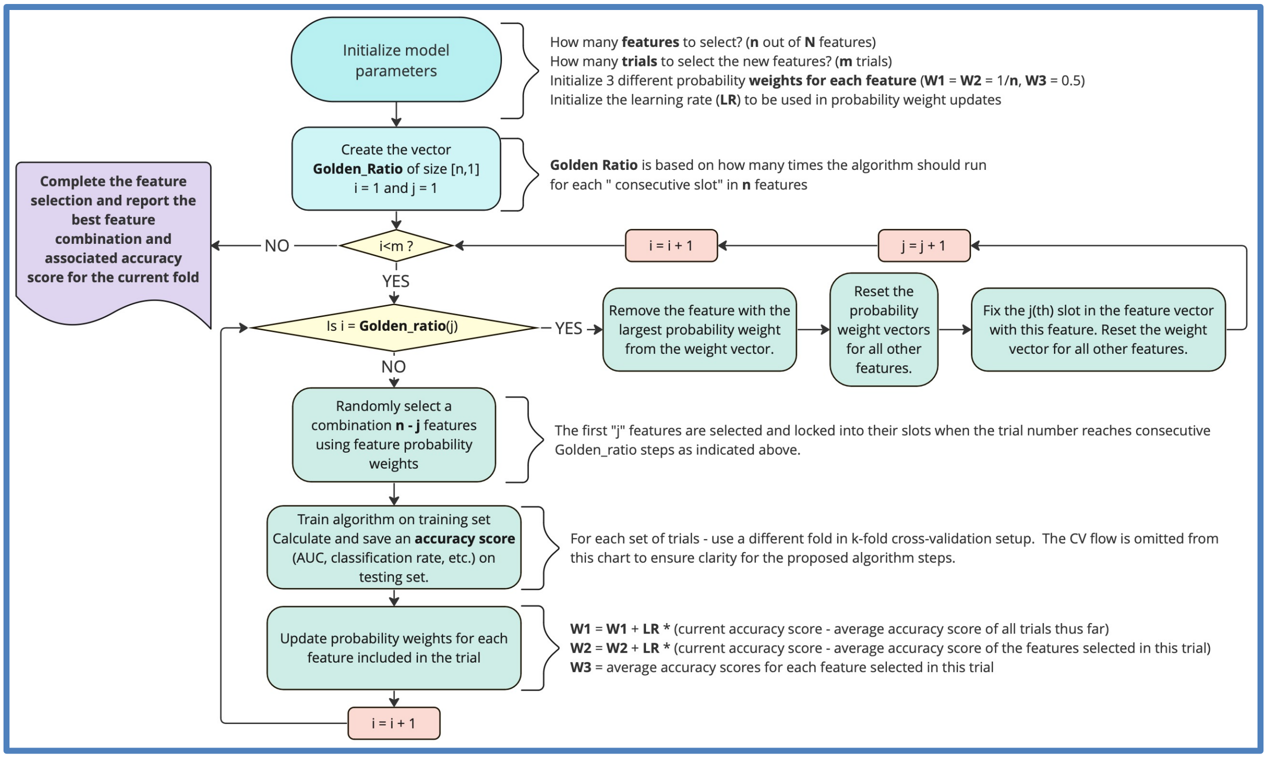


Figure 3: Probability weighted feature selection (PWFS) algorithm to achieve the best empirical performance on medium sized datasets

The PWFS algorithm was coupled with a Random Forest classifier (Pal, 2005) for binary classification of whether the student GPA was higher or lower than 3.33. A random forest binary classifier combines the predictions of multiple decision trees to make a final decision between two binary outcomes. Each decision tree in the forest makes an independent prediction, and majority voting decides the outcome. The ultimate decision is binary and categorical (instead of predicting the numerical GPA value) and the choice of the learning algorithm was finalized after an empirical search of different algorithms and topologies including standard logistical regression and multilayer perceptron (i.e, neural networks). The proposed feature selection method was developed due to low performance when all survey questions were equally represented at the input of the classification algorithm.

Results

The results are presented in three subsections to help answer the individual research questions identified in the previous sections.

Can a predictive model be trained on the survey responses with sufficient accuracy compared to the baseline (in this case 50% for a binary classification) in classifying student GPA

groups as TRUE (higher than 3.3 GPA) or FALSE (lower than 3.3 GPA)?

In order to answer this question, the machine learning model consisting of PWFS feature selection algorithm and a random forest binary classifier were trained and tested on the main partition which includes a total of 1098 survey responses (both PRE and POST) from 549 unique students in each group. We applied 10-fold cross-validation and averaged the overall responses in a singular confusion matrix as shown below:

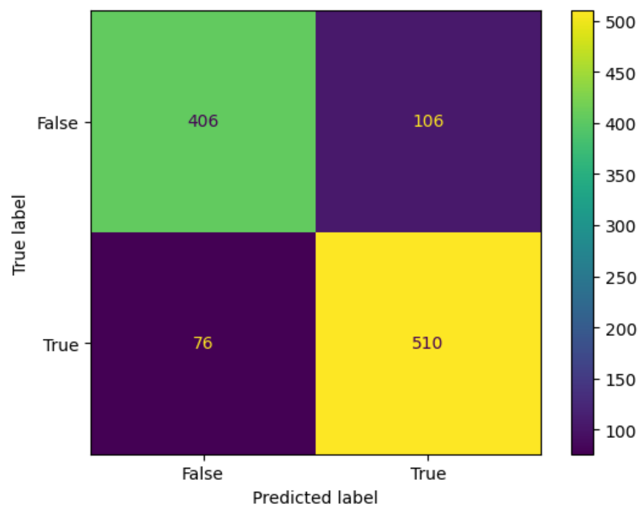


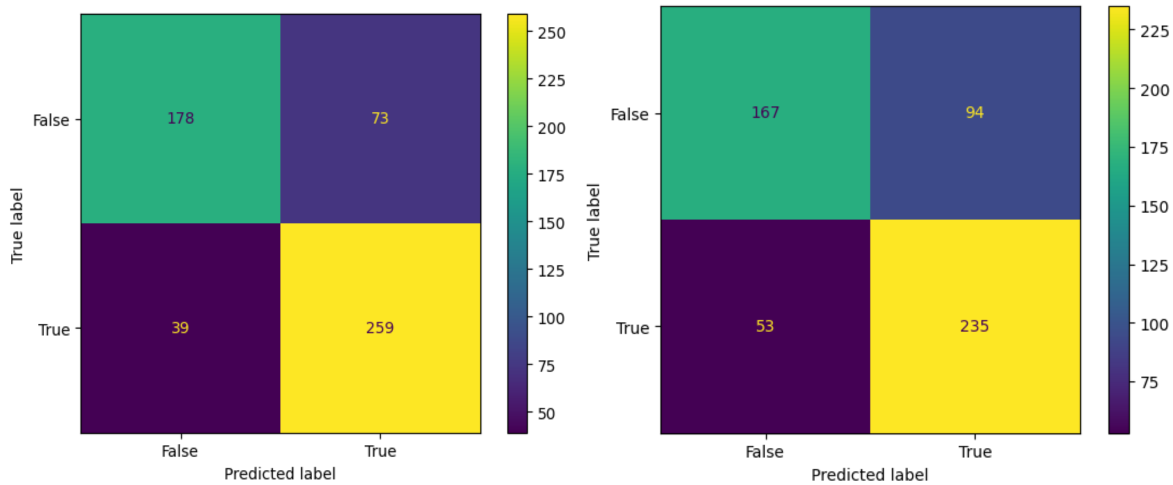
Figure 4: Confusion matrix where the model was trained and validated on the main partition of the dataset

The average accuracy can be found by adding the diagonals (which indicate correctly classified samples) and dividing by the total sum (which also includes incorrectly classified samples). In this case the average accuracy is ~83.4% with a corresponding error rate of ~16.6%. These results not only significantly outperform our latest findings presented in 2023 ASEE by approximately a 55% reduction in error rate but also answers the first research question affirmatively with a 33.4% improvement over "chance prediction" where 4 in 5 students are labeled in the correct GPA category.

If the answer is yes to question 1, does the model trained only on pre-intervention action state surveys have quantifiable levels of difference in accuracy when tested on the post-intervention responses of students?

In order to study this objective, we used the two partitions of the main dataset labeled PRE and POST (and shown in Tables 3 and 4 previously). First, the model was trained only on PRE partition (once again using 10-fold cross validation as in the previous section). The trained model was then tested on the POST partition (with no need for cross-validation as the entire POST partition is now included in the test data). The results are reported in the two confusion matrices as shown below in figure 5.

Once again, the average accuracy for each configuration can be found by adding the diagonals (which indicate correctly classified samples) and dividing by the total sum (which also includes



(a) ***(b)***

Figure 5: Confusion matrices where the model was trained and validated only on the PRE partition of the dataset (a) and later tested on the POST partition of the dataset (b)

incorrectly classified samples). In this case the average accuracy for the model trained and tested on the PRE partition is ~79.6% (a), whereas the average accuracy for when it's tested on the POST partition is ~73.2% (b). More importantly, the corresponding error rates are ~20.4% (a) and ~26.8% (b) which represent a significant ~32% increase in error rate when the model trained on PRE data was tested on POST data. It is important to note at this stage that even though the survey responses are unique, they still belong to the same students included in both datasets. This indicates that there is in fact a quantifiable level of difference in accuracy when the same model is used to predict PRE and POST survey academic performance.

If the answer is yes to question 2, is this difference in accuracy reflected in explainable and modest changes in false-positive ratios between the models trained and tested on different populations?

Finally, to understand the "direction" of change, we explore the false-positive rates for the two scenarios. False positive rate is defined as the ratio of FALSE samples incorrectly classified as TRUE and the overall number of FALSE samples. The false positive rate for the model trained and tested on PRE partition is 73 / (178+73) ~ 29% (a). The false positive rate for the same model when tested on POST partition is 94 / (167+94) ~ 36% (b).

This represents an almost 25% increase in the false-positive rate for the model in the second scenario. In other words, when looking at the population, the model trained on survey responses recorded before the intervention "mistakenly" thinks that the same student who took the survey after the intervention has a higher GPA when in reality it is not the case. This can be inferred as the intervention creating study habits in students more representative of high GPAs since simply comparing GPAs is not possible due to a range of reasons (including but not limited to advancing in seniority and taking more difficult classes).

Conclusions and Future Work

This study accomplished two major objectives in exploring how action-state orientation impacts student performance and whether any improvement in study habits due to in-class interventions can be quantified using state-of-the-art machine learning methods.

A random forest classifier using our proposed PWFS feature selection method was able to accomplish 83.4% classification rate on the question of whether a student has higher or lower GPA compared to the threshold value of 3.33 across a wide-range of student cohorts from multiple semesters and a period of two years. This indicates that there is relevant information in these responses which can be used to infer the likelihood of a student being successful academically. Our hypothesis for the second objective was to see whether this information can be used to quantify any "potential" improvement in academic performance. It is important to note that scientifically conducting an "improvement" study on an objective metric such as GPA is practically impossible. The students take before-and-after surveys either in the same semester (in which case their GPA will be the same) or in subsequent semesters where more or less challenging course work will have a direct impact on their GPA. This would require an unobtainable number of survey responses to statistically significantly separate and highlight the impact of such interventions.

Thanks to a novel approach in which an algorithm is trained only on PRE-intervention responses and tested on POST-intervention responses, we were able to demonstrate a near 25% increase in associated false-positive rates which indicate that the ML model mistakenly believes a student is successful based on the study habits it learned from prior training which leads to higher GPA. For instance, a student can be classified as having a high GPA after going through an intervention even when their GPA has either not changed or even potentially gotten lower. Since testing for what has not happened yet is not an option, the ML model provides a window into "potential" improvements in student performance due to changing study habits.

Future work will include individualized changes in model performance to identify whether some students are more likely to benefit from any personalized intervention. We will continue to add new survey data for the remainder of the project lifetime which will only help improve prediction performance while providing a clearer window into the yet-to-be-realized benefits of such interventions.

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

Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, *26*(1), 217-222.

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

Uysal, I., Spector, P. E., Ferekides, C. S., Ayanoglu, M. B., & Elashmawy, R. (2023, June). Predicting Academic Performance for Pre/Post-Intervention on Action-State Orientation Surveys. In 2023 ASEE Annual Conference & Exposition.