Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Application of Neighborhood Components Analysis to Process and Survey Data to Predict Student Learning of S

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Lu, Yikai	University of Notre Dame	
Ober, Teresa	University of Notre Dame	0000-0001-9698-9543
Liu, Cheng	University of Notre Dame	0000-0002-2787-1653
Cheng, Ying	University of Notre Dame	0000-0002-8654-1443

Publication/Completion Date—(if *In Press,* enter year accepted or completed) August 17, 2022

Check type of content being submitted and complete one of the following in the box below:

- ☐ If article: Name of journal, volume, and issue number if available
- ☑ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

2022 International Conference on Advanced Learning Technologies (ICALT); July 1- 4, 2022; Hybrid

DOI or URL to published work (if available) 10.1109/ICALT55010.2022.00051

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scie through **[Grant number]** R305A1802 to **Institution]** University of N. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

Application of Neighborhood Components Analysis to Process and Survey Data to Predict Student Learning of Statistics

Yikai Lu
Department of Psychology
University of Notre Dame
Notre Dame, IN, USA
ylu22@nd.edu

Teresa M. Ober
Department of Psychology
University of Notre Dame
Notre Dame, IN, USA
tober@nd.edu

Cheng Liu
Department of Psychology
University of Notre Dame
Notre Dame, IN, USA
cliu7@nd.edu

Ying Cheng
Department of Psychology
University of Notre Dame
Notre Dame, IN, USA
ycheng4@nd.edu

***Abstract*—Machine learning methods for predictive analytics have great potential for uncovering trends in educational data. However, simple linear models still appear to be most widely used, in part, because of their interpretability. This study aims to address the issues of interpretability of complex machine learning classifiers by conducting feature extraction by neighborhood components analysis (NCA). Our dataset comprises 287 features from both process data indicators (i.e., derived from log data of an online statistics learning platform) and self-report data from high school students enrolled in Advanced Placement (AP) Statistics (*N*=733). As a label for prediction, we use students' scores on the AP Statistics exam. We evaluated the performance of machine learning classifiers with a given feature extraction method by evaluation criteria including F1 scores, the area under the receiver operating characteristic curve (AUC), and Cohen's Kappas. We find that NCA effectively reduces the dimensionality of training datasets, stabilizes machine learning predictions, and produces interpretable scores. However, interpreting the NCA weights of features, while feasible, is not very straightforward compared to linear regression. Future research should consider developing guidelines to interpret NCA weights.**

***Index Terms*—predictive analytics, machine learning, neighborhood components analysis, interpretability, Advanced Placement**

I. Introduction

The ability to predict factors that place students at risk for underachievement in an early stage of a course given known variables (e.g., demographic factors, past academic history, etc.) is often a goal of learning analytics research [1]. Predictive analytic modeling has been proposed as one means by which to use different sources of educational data to predict students' academic success [2, 3]. One challenge to developing better predictive models of student learning is to balance both accuracy and interpretability. An ideal predictive model is one that is both highly interepretable yet also retains good predictive accuracy. If teachers and students can understand the results of a predictive model in a way that informs them about student learning, they are more likely to benefit from them. Though approaches and applications for predictive analytics modeling may differ, striving for interpretability and predictive accuracy should be a goal of learning analytics research.

This work was supported by a National Science Foundation CAREER (Grant #DRL-1350787) and an Institute for Education Sciences (Grant #R305A180269) awarded to Dr. Ying Cheng.

A. Features Used for Predictive Analytic Modeling

The qualities of features used in predictive analytic modeling is an important consideration that affects the interpretability and accuracy of predictive models. Increasingly, process data from students' completion of online learning activities has been used to predict student modeling. Process data (also referred to as "digital log data" or "trace data") is generated as a user navigates an online Learning Management System (LMS). Particularly given that it can be gathered in a manner unobtrusive to the learning activities, process data appears to be one promising source of information for predictive modeling of students' academic success. A variety of features gathered unobtrusively from LMS process data have been used in student predictive modeling, for example, test completion status [4], student demographic factors [5], clickstream data [6, 7], progress rates of assignments [8], and past course success [6, 9, 10]. With few exceptions [5, 10, 11], current learning analytics research typically uses data gathered from only one type of data source as input to predict learning outcomes. One possible reason for this is that it is not easy to interpret models that leverage data from multiple sources including surveys, click-stream, and assessments.

B. Methods for Predictive Analytic Modeling

Aside from the types and variety of features used in predictive modeling, the analytic methods used can also affect model intepretability and accuracy. Linear regression appears to be among the most commonly used techniques in predictive analytic modeling to classify student performance [12], likely due to the relative simplicity of interpretation. However, simple linear models such as logistic regression, which are more interpretable, might not handle complex or high-dimensional data sets very well. By contrast, more sophisticated machine learning methods (e.g., support vector machine or deep neural networks) involve nonlinear transformations which may provide more accurate predictions yet may also complicate the interpretability of the results. Given the importance of interpretability in educational research, we would like to maintain both goals of interpretability and prediction accuracy.

C. Feature Selection and Feature Extraction

There are typically two kinds of data preprocessing methods used to handle high-dimensional data in constructing predictive models: feature selection and feature extraction. Feature selection reduces the dimension of input data by removing trivial features beforehand (i.e., selecting features with larger assigned importance scores). Feature extraction has an advantage over feature selection when linear transformation is used for feature extraction. Feature extraction simplifies the data by transforming it from a high-dimensional space into a low-dimensional space. The results are easier to comprehend when linear transformation is adopted for feature extraction because the weights can be interpreted as they would be a linear regression model. Thus, we can use feature extraction methods for data preprocessing to improve the interpretability of predictive models.

1) Different Methods for Feature Extraction: To improve model interpretability without drastically compromising prediction accuracy, unsupervised dimensionality reduction and a distance metric learning method could be used to determine which and to what extent certain features would contribute the most predictive potential to a certain outcome. One useful yet under-examined distance metric learning method, neighborhood components analysis (NCA) [13], has the potential to derive a metric useful to both students and teachers to understand student learning.

NCA is a supervised learning technique that tries to find a linear transformation of the training dataset that predicts the outcome well with k-nearest neighbors (kNN) [13]. Technically speaking, NCA maximizes "a stochastic variant of the leave-one-out kNN score on the training set." [13] Here, "a stochastic variant" means that the probability density function uses a softmax function given Euclidean distances in the transformed space, rather than the actual probability of a certain class in kNN. More specifically, NCA maximizes the objective function of the expected number of data points correctly classified:

$$f(A) = \sum_{i} \sum_{j \in C_i} \frac{\exp(-||\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_j||^2)}{\sum_{k \neq i} \exp(-||\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_k||^2)}, \quad (1)$$

where $\mathbf{x}_i \in \mathbb{R}^m$ is a data point with m features, $j \in C_i$ denotes the points in the same class i, and $\mathbf{A} \in \mathbb{R}^{d \times m}$ is a transformation matrix where d is the number of dimensions in an output vector.

2) Presumed Advantages of NCA over Other Methods: There are several advantages of using NCA over other methods for feature extraction. First, NCA over other methods is that NCA relies on fewer assumptions compared to traditional methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) [14]. PCA and LDA extract information only from covariance structures of data, which typically requires normality assumption. However, NCA does not need such assumption, since it solely relies on kNN. Thus, NCA can extract complex patterns in data that PCA or LDA cannot [13]. Second, NCA provides interpretable information on which features are especially important in predicting the outcome. Linearly transformed values of the training dataset, which we call NCA scores, are interpretable because similar NCA scores suggest similar outcomes, as the objective function optimizes k-nearest neighbors of data points based on the label. In NCA, the dimension of the transformation matrix, $\mathbf{A}$, implies the complexity of association between features and label. For example, when we only need one row in $\mathbf{A}$ to achieve a desired accuracy, it suggests that the association between features and label is one dimensional. Third, NCA can be coupled with a more sophisticated classifier by taking its output as input to the classifier.

II. Current Study Aims

We used process data from an online statistics learning platform, self-report survey data, and assignment scores as features. Scores from a standardized high-stakes exam, the Advanced Placement (AP) Statistics exam, were used as a label for prediction. Though past research has attempted to predict student performance based on process data from an LMS, few studies have attempted to use both features derived from process data in addition to those from conventional psychological and educational measures. Using data from multiple sources, we sought to illustrate how NCA can be applied in predictive modeling in the context of educational research. Our work makes contributions in at least two aspects. First, we demonstrate the potential for analyzing high-dimensional data consisting of features derived from process data as well as psychological survey responses and educational assessments. Second, by applying NCA to conduct feature extraction and prediction simultaneously, we develop a predictive model that maintains high prediction accuracy yet still parsimonious, despite the high-dimensionality of the data.

III. Data and Analyses

A. Data

Subjects (N=733) included high school students enrolled in AP Statistics courses during the 2018-2019 and 2019-2020 academic years. They were between the ages of 14–18 years of age ($M = 16.76, SD = 0.87$). There were slightly more female students than male students (Female = 53.7%, Male = 47.3%). Of those subjects who reported their racial/ethnic identity (71.9%), most identified as White/European American (62.6%), Asian/Asian American (15.6%), Black/African American (5.3%), Mexican American (2.3%), Other Hispanic or Latino/Latina (2.1%), with the remainder reporting as multiracial or other (12.1%).

Subjects used an online assessment and learning platform specifically designed for AP Statistics courses. In addition to assignment and score management, the platform uses computerized adaptive testing, which applies item response theory

(IRT) to improve person ability estimates by administering optimal items based on students' previous responses. The dataset contains several kinds of features, including demographic information, as well as students' responses to surveys measuring students' math attitudes [15], engagement in the course (i.e., micro-engagement) [16] and school (i.e., macro-engagement) [17], personality (Big Five Inventory; BFI-2) [18], and academic procrastination [19], in addition to assignment scores (i.e., person ability estimates by IRT), and the number of times a student checked their assignment score reports (i.e., results-checking counts) derived from the process data. We use one-hot encoding (i.e., treating each value of a categorical variable as a dummy variable which represents 0/1) for survey question items such that more than one option can be selected. The resulting training data matrix has 287 features consisting of all these kinds of features. As noted previously, we use the final AP Statistics exam score as a label for prediction.

B. Data Preprocessing

1) Class Imbalance Handling: The distribution of the label classes consists of 59 people with a score of 1, 110 people with a score of 2, 180 people with a score of 3, 143 people with a score of 4, 150 people with a score of 5, and 91 people with missing values. Missing values most likely reflected the student's decision not to take the AP exam. Receiving a score of 1 or 2 on the AP statistics exam indicates a student would not be eligible to receive college credit given the minimum accepted score is reported to be 3 or 4 for most colleges [20]. Furthermore, there were few students who got either 1 or 2 on AP Statistics exam relative to those who received a score of 3, 4, or 5. Therefore, we combined students who received a 1 or 2 into one class prior to data analysis. After recoding, class imbalance was no longer an issue and thus no sampling method is used.

2) Missing Data Handling: Missing values were imputed for features. We used kNN imputation for each feature because it is useful for handling various kinds of missing data [6]. Due to the variety of the variable types in the data, variables were standardized before kNN imputation was computed so that each feature had the same scale. The kNN imputation was conducted by scikit-learn with the default options [21].

C. Feature Extraction

We used NCA to transform the training dataset into a single linear scale (or more components, if needed) using AP exam scores as a criterion. Figure 1 shows the validation curve with kNN after data were transformed by NCA. We repeated 10-fold cross validation five times. Preliminary analyses (explained further in the "Results" section) showed that the one component solution performed comparably to the two component solution. For the sake of model parsimony, we kept one component from NCA.

D. Machine Learning Classifiers

After data preparation was complete, we then entered the raw data or NCA scores into machine learning models to predict AP exam scores. The python library scikit-learn was used to run popular classifiers, which included logistic regression (LR), naïve Bayes (NB), support vector machine (SVM), decision tree (DT), kNN, random forest (RF), multilayer perceptron (MLP), and AdaBoost (AB) [21]. We also used a stratified dummy classifier (Dummy) as the baseline which is a classifier that randomly selects a label class based on class prior probabilities. Different criteria (i.e., precision, recall, and F1 scores) were evaluated using an average weighted by support. Since the F1 score is the harmonic mean of precision and recall, we used it as a threshold-sensitive measure of accuracy. AUC scores, which are calculated by the average of all possible pairwise combinations of classes, were used to evaluate model efficacy. Weighted Cohen's kappa was used as a primary index to evaluate the performance because if considers the distance between of the label classes (recoded AP scores) by quadratic weights.

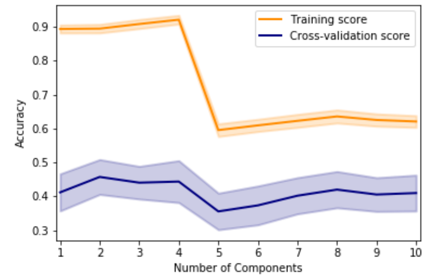

Fig. 1. Validation curve with kNN after data were transformed by NCA.

IV. Results

A. Evaluation for Class Prediction

Table I shows the prediction results. Since the results of the Dummy condition are expected to be the same, we omit the Dummy condition for NCA conditions. For the raw data condition, random forest performed the best, producing the highest F1 score, Cohen's kappa, and AUC score. SVC performed second best. These results show that the two models can be used to predict AP Statistics scores from raw data. However, both random forest and SVC are difficult to interpret, given that random forest is an ensemble method that uses decision trees and SVC uses kernalization which would transform raw data in uninterpretable ways.

NCA often outperforms raw data conditions or at least achieves comparable results. Notably, with NCA, more interpretable models such as LR, kNN, and NB performed much better. It is also more stable than just using the raw data in terms of Cohen's kappas. In particular, NCA led to Cohen's kappas typically around or above 0.6, suggesting moderate and substantial agreement [22]. For instance, using NCA with kNN achieved Cohen's kappa of 57.2% on average, while using raw data achieved 39.1%, suggesting that similar NCA scores will predict similar outcomes.

TABLE I
RESULTS OF PREDICTIVE MODELS WITH DIFFERENT FEATURE EXTRACTION METHODS EVALUATED WITH VARIOUS CRITERIA (NUMBERS ARE IN PERCENTAGE)

	LR	kNN	DT	NB	SVC	RF	AB	MLP	Dummy
Raw Data									
Accuracy	42.5 (5.48)	36.6 (5.12)	47.8 (5.33)	35.1 (4.47)	49.3 (6.34)	49.9 (5.86)	45.6 (5.6)	41.9 (7.44)	20.9 (4.81)
Precision	43.3 (5.42)	37.0 (6.43)	47.1 (6.22)	33.9 (11.61)	49.5 (6.66)	51.1 (6.03)	47.7 (6.34)	41.8 (7.61)	21.1 (5.03)
Recall	42.5 (5.48)	36.6 (5.12)	47.8 (5.33)	35.1 (4.47)	49.3 (6.34)	49.9 (5.86)	45.6 (5.6)	41.9 (7.44)	20.9 (4.81)
F1 Score	42.1 (5.39)	33.9 (5.45)	45.1 (5.05)	27.6 (4.7)	48.7 (6.4)	49.2 (5.71)	42.8 (5.39)	41.2 (7.33)	20.7 (4.77)
AUC	73.9 (3.13)	67.5 (4.2)	77.2 (3.62)	71.9 (4.23)	79.8 (3.18)	79.5 (3.79)	74.8 (3.02)	73.3 (4.99)	50.7 (2.77)
Cohen's kappa	52.9 (10.09)	39.1 (10.45)	62.7 (8.31)	47.8 (8.78)	62.0 (8.03)	63.9 (8.88)	58.1 (7.81)	54.3 (10.06)	1.7 (10.63)
NCA									
Accuracy	45.2 (5.33)	41.1 (5.14)	45.3 (4.76)	45.4 (5.21)	45.7 (5.17)	45.4 (6.15)	42.2 (7.06)	45.1 (5.03)	
Precision	46.3 (5.74)	41.0 (4.99)	47.5 (5.37)	46.7 (5.62)	48.2 (5.1)	47.0 (6.68)	37.6 (7.48)	45.2 (6.66)	
Recall	45.2 (5.33)	41.1 (5.14)	45.3 (4.76)	45.4 (5.21)	45.7 (5.17)	45.4 (6.15)	42.2 (7.06)	45.1 (5.03)	
F1 Score	44.2 (5.25)	40.5 (4.94)	45.0 (4.72)	44.7 (5.18)	45.3 (5.17)	44.9 (6.29)	38.8 (7.18)	43.0 (5.5)	
AUC	75.8 (3.2)	66.1 (3.78)	70.9 (3.87)	76.7 (3.37)	76.4 (3.76)	71.0 (4.6)	72.9 (3.57)	76.4 (3.25)	
Cohen's kappa	63.2 (7.03)	57.2 (8.64)	61.8 (7.71)	64.2 (6.91)	64.3 (8.17)	63.0 (8.04)	58.8 (8.36)	63.3 (6.69)	

B. Interpretations for Class Prediction

NCA performs better in many evaluation criteria with only one component, so we focus on interpretation of this model. Table II shows the top ten most important features in the NCA component, as determined by weights in $\mathbf{A}$ with the largest absolute values. We can infer that certain variables have large influence on students' AP exam scores, for example whether or not students dropped the class, their assignment submissions or assignment scores, as well as their predictions of their AP exam score (indicative of students' self-efficacy).

V. DISCUSSION

We found that NCA effectively reduces the dimension of the training dataset while maintaining - or even improving - prediction performance based on various criteria over raw data. NCA had more stable Cohen's kappa values, especially when used with the kNN classifier. These results suggest having similar feature values does not effectively predict similar outcomes. By contrast, having similar NCA scores does effectively predict similar outcomes. As such, there is an advantage in using the NCA scores. Further, one can interpret NCA scores by looking at the weights and values of the features. These findings suggest that NCA has great promises in exploring complex feature-label association.

A. Future Work and Limitations

Despite the potential applications of this approach based on our findings, there are several limitations in generalizing the findings beyond the present study context. First, although the number of the features (287) is large, the sample size ($N = 733$) is not large enough to generalize the NCA weights. Currently, the standard error of the estimated weights is indeterminant and no statistical inference method is available to investigate whether this difference is significant. To mitigate this issue, having a larger sample size is especially important.

Second, during our data collection, COVID-19 likely affected students' behavior and outcomes. Our own findings suggest the cohort of students enrolled in the 2019-2020 academic year showed large behavioral changes. Additionally, the AP Statistics exam format and content changed during the 2019-2020 academic year as a result of COVID-19.

Our analysis used one NCA component, which is relatively easy to interpret. Future research should investigate datasets that require more than one component to achieve good performance. Another limitation of the methods used in the present study is that NCA weights could be sensitive to how features are coded. Future work should attempt to investigate NCA alongside other feature extraction methods (e.g., PCA, LDA) to determine whether the presumed benefits of using NCA hold or could explore the possibility of combining explanatory methods such as factor analysis or structural equation modeling with NCA.

ACKNOWLEDGMENT

We would like to thank the high school teachers and students who participated in this project.

REFERENCES

[1] Y. Cui, F. Chen, S. Ali, and Y. Fan, "Predictive analytic models of student success in higher education," *Information and Learning Science*, vol. 120, no. 3/4, pp. 208–227, 2019.

[2] K. Chrysafiadi and M. Virvou, "Student modeling approaches: A literature review for the last decade," *Expert Systems with Applications*, vol. 40, no. 11, pp. 4715–4729, 2013.

[3] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020.

[4] J. Figueroa-Cañas and T. Sancho-Vinuesa, "Early prediction of dropout and final exam performance in an online statistics course," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 15, no. 2, pp. 86–94, 2020.

[5] M. Tan and P. Shao, "Prediction of student dropout in e-learning program through the use of machine learning

TABLE II
TOP TEN NCA FEATURES (POSITIVE DIRECTION SUGGESTS BETTER A SCORE)

Rank	Explanation for Feature	Direction	Weight
1	Student dropped the class or not	-	1433.213
2	Mock AP Score	+	637.157
3	AP Statistics score predicted by student	+	367.158
4	The total length of time for all problem-solving sessions	-	362.046
5	My teacher cares about how much I learn	+	277.703
6	An assessment group from a relatively poor region	-	276.639
7	What is your racial or ethnic identification? - "Black"	-	276.614
8	Student score of Assignment 4	+	270.154
9	Student has taken Assignment 3	+	268.148
10	Assignment 3 Start to Finish Duration	-	257.335

method." *International Journal of Emerging Technologies in Learning*, vol. 10, no. 1, 2015.

[6] F. Chen and Y. Cui, "Utilizing student time series behaviour in learning management systems for early prediction of course performance," vol. 7, no. 2, pp. 1–17, 2020, number: 2.

[7] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, "Machine learning approaches to predict learning outcomes in Massive open online courses," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 713–720, iSSN: 2161-4407.

[8] L. Faucon, J. K. Olsen, S. Haklev, and P. Dillenbourg, "Real-Time Prediction of Students' Activity Progress and Completion Rates," *Journal of Learning Analytics*, vol. 7, no. 2, pp. 18–44, Sep. 2020, number: 2.

[9] A. Tekin, "Early prediction of students' grade point averages at graduation: A data mining approach," *Eurasian Journal of Educational Research*, vol. 54, pp. 207–226, 2014.

[10] K. E. Arnold and M. D. Pistilli, "Course signals at purdue: using learning analytics to increase student success," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*. ACM Press, 2012, p. 267.

[11] C. E. López Guarín, E. L. Guzmán, and F. A. González, "A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 10, no. 3, pp. 119–125, Aug. 2015, conference Name: IEEE Revista Iberoamericana de Tecnologias del Aprendizaje.

[12] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Applied Sciences*, vol. 11, no. 1, p. 237, 2021.

[13] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Proceedings of the 17th International Conference on Neural Information Processing Systems*, ser. NIPS'04. MIT Press, 2004, pp. 513–520.

[14] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[15] S. Y. Lim and E. Chapman, "Development of a short form of the attitudes toward mathematics inventory," *Educational Studies in Mathematics*, vol. 82, no. 1, pp. 145–164, 2013.

[16] B. M. Whitney, Y. Cheng, A. S. Brodersen, and M. R. Hong, "The Scale of Student Engagement in Statistics: Development and Initial Validation," *Journal of Psychoeducational Assessment*, vol. 37, no. 5, pp. 553–565, Aug. 2019, publisher: SAGE Publications Inc.

[17] S. R. Hart, K. Stewart, and S. R. Jimerson, "The Student Engagement in Schools Questionnaire (SESQ) and the Teacher Engagement Form-New (TERF-N): Examining the preliminary evidence," *Contemporary School Psychology*, vol. 15, pp. 67–79, 2011, place: US Publisher: California Assn of School Psychologists.

[18] C. J. Soto and O. P. John, "The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power," *Journal of Personality and Social Psychology*, vol. 113, no. 1, pp. 117–143, 2017, place: US Publisher: American Psychological Association.

[19] M. E. Aitken, *A personality profile of the college student procrastinator*. University of Pittsburgh, 1982.

[20] CollegeBoard, "Ap credit policy search," https://apstudents.collegeboard.org/getting-credit-placement/search-policies/course/31, 2022.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[22] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," vol. 33, no. 1, pp. 159–174, 1977, publisher: [Wiley, International Biometric Society].