# Clustering Students in ASSISTments: Exploring System- and School-Level Traits to Advance Personalization

Seth Adjei, Korinn Ostrow, Erik Erickson, Neil Heffernan
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
{saadjei, ksostrow, eerickson, nth}@wpi.edu

## ABSTRACT
Few attempts have been made to create student models that cluster student and school level traits as a means to design personalized learning interventions. In the present work, data from ASSISTments was enriched with publicly available school level data and K-Means clustering was employed. Results revealed the importance of school locale, measures of district wealth, and system interaction patterns as potential foci for personalization. Clusters were then applied to a test set of held out data and cluster assignments were used to help predict end-of-year standardized mathematics test scores. Findings suggest that while cluster interpretations were not generalizable to held out data, clustering was generally helpful in predicting standardized test scores.

## Keywords
K-Means Clustering, Student-System Interactions, School Level Characteristics, Standardized Tests, Ensembled Prediction Model.

## 1. INTRODUCTION
The focus of research using vast educational data often lends itself to the development of learner models, or various sophisticated predictive models that help to pinpoint when and how learning occurs on a personalized level. Popular approaches include Bayesian Networks (i.e., Bayesian Knowledge Tracing) [3], Performance Factors Analysis [6], and Neural Networks (i.e., Deep Learning) [4]. However, it is valuable to ask if simpler models built to leverage student, school, and district level data can be useful in establishing learner profiles.

The use of clustering to group similar students within various types of online learning environments has typically been a successful endeavor [1, 2, 7, 8]. The present work seeks to balance the complexity of working with high volumes of educational data and building simple predictive learner models through clustering by answering the following research questions:

1. Are there distinct types of learners within ASSISTments [5] that can be identified by clustering student, school, and district level characteristics and measures of student/system interaction?
2. What student types are defined via cluster interpretation? Do interpretations generalize to unseen data?
3. Can clusters help predict significant differences in end-of-year test scores?

## 2. METHODOLOGY
The present work assessed log files from students in the state of Maine working in ASSISTments [5], an online learning system focused on middle school mathematics, during the 2014-2015 academic year. This data was extended by merging additional school and district level data from the Common Core of Data supported by the NCES and IES (https://nces.ed.gov/ccd/). Students' scores on the standardized, end-of-year TerraNova mathematics test were also included in the dataset.

For each student, the dataset contained averages for the following student/system interaction features: problem count, time spent on problems, percent correct across assignments, hints used per problem, number of problems per assignment for which hints were used, and assignment completion rate. Additionally, each student's data included continuous measures retrieved from the NCES/IES data (i.e., the percentage of students in the school eligible for free or reduced lunch) as well as one-hot encoded forms of categorical features like school locale. The cleaned dataset represented 1,557 unique students from 21 schools, with 171,983 unique student/assignment pairs stemming from 35,127 assignments. Each observation or row represented the overall performance and characteristics of a single student and their school or district. De-identified data is available at tiny.cc/EDM2017Clustering for further reference.

The modeling approach used in the present work was adapted from that in [1]. An initial 70% of the data was randomly selected to form the training set. The training set was used for initial K-Means clustering and cluster interpretation. The K-Means algorithm was sourced from R's statistics package, implementing Euclidean distance as the default distance measure. The remaining 30% of the data was used to form the test set. The test set was used to build models predicting TerraNova scores. First, predictions were made to assign students in the test set to a cluster. Following student assignment, clusters were reinterpreted to verify whether trained interpretations generalized to unseen data. Cluster membership was then used to help predict TerraNova scores alongside student-system interaction features using cluster-specific stepwise linear regressions. These regression models were then ensembled and measures of model accuracy were compared to a traditional approach where $K = 1$.

## 3. TRAINING
In order to determine the optimal value for K, 10-fold cross validation was implemented on the training set to build scree plots. To determine the most appropriate value from this set, the mean and median of optimal K values across folds were considered ($M = 4.1$, Med. = 4). As such, four clusters were forced using K-Means on the training data. The four resulting clusters were characteristic of unique types of students, ultimately labeled as "proficient," "struggling," "learning," and "gaming." Graphics and additional information on cluster characteristics are available at tiny.cc/EDM2017Clustering for further reference.

**Table 1. Coefficients, Standard Errors, and Model Statistics per cluster on test set data when K=1 and K=4.**

| | *K = 1* | | *K = 4* | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *1 (n = 442)* | | *1 (n=127)* | | *2 (n=160)* | | *3 (n=124)* | | *4 (n=31)* | |
| **IVs** | *b* | *SE* | *b* | *SE* | *b* | *SE* | *b* | *SE* | *b* | *SE* |
| **Intercept** | 631.94*** | 20.37 | 712.95*** | 51.78 | 504.41*** | 30.36 | 567.63*** | 34.66 | *680.14**** | *63.13* |
| Percent Correct | 110.66*** | 22.76 | 81.95 | 61.70 | 268.30*** | 33.92 | 131.02*** | 35.16 | 18.73 | 68.74 |
| Ave. Time | -0.08** | 0.03 | -0.10 | 0.07 | 0.01 | 0.04 | 0.09 | 0.06 | -0.09 | 0.09 |
| Completed | 0.35 | 12.13 | -63.05 | 39.89 | 8.47 | 15.55 | 22.10 | 18.68 | -18.80 | 34.25 |
| Total Hints | 1.73 | 2.69 | 7.01 | 6.08 | 8.00* | 3.66 | -38.84*** | 8.02 | -52.73* | 19.80 |
| Hint Instances | -0.11 | 3.53 | -9.34 | 11.25 | -4.13 | 4.13 | 49.75*** | 9.68 | 71.09** | 24.23 |
| **Model Stats** | | | | | | | | | | |
| F (DF) | 17.55*** (5, 436) | | 1.30 (5, 121) | | 22.87*** (5, 154) | | 8.18*** (5, 118) | | 2.00 (5, 25) | |
| $R^2$ (Adj. $R^2$) | 0.168 (0.158) | | 0.051 (0.012) | | 0.426 (0.408) | | 0.257 (0.226) | | 0.286 (0.143) | |

## 4. TESTING & MODEL EVALUATION

Using the remaining 30% of the data that had been held out from the training set, student, school, and district level features (excluding TerraNova test score) were used to predict student assignment to one of the four clusters developed in training. Following student assignment, clusters were interpreted to verify whether initial cluster labels generalized to this unseen data. Cluster characteristics varied for the test set, suggesting that cluster interpretations did not generalize. Graphics and additional information on cluster characteristics are available at tiny.cc/EDM2017Clustering for further reference.

Cluster membership was then used to help predict TerraNova scores alongside student/system interaction features using cluster-specific stepwise linear regressions. Following the ensembling approach used in [7], separate regression models were built for each cluster before being ensembled to form a prediction model. Cluster models helped to depict the relative importance of student/system interaction features in the prediction of TerraNova scores for each value of K, as shown in Table 1. Variability in feature significance was observed across clusters. An alternative prediction model was constructed using the full dataset (essentially, K=1) in order to compare the accuracy of ensembled cluster models to an unclustered baseline. Table 1 presents unstandardized beta coefficients, standard errors, significance values, and overall model statistics across clusters and values of K, and reveals that cluster assignment was sometimes significant in predicting TerraNova scores.

In terms of prediction model accuracy, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were both lowest when K=4 (23.27 and 30.32, respectively, compared to 25.88 and 33.44 when K=1). Additionally, the difference between MAE and RMSE was lower when K=4 (7.05 compared to 7.56), suggesting that the variance in individual prediction errors decreases as K increases. Variance explained, as measured by $R^2$, was also higher when K=4, suggesting that the ensembled model was a stronger option than grouping all data together into a single cluster.

## 5. DISCUSSION

Results of our clustering exploration revealed that there are distinct types of learners within ASSISTments that can be identified by using K-Means to cluster student, school, and district level characteristics and measures of student/system interaction. Results suggested that clusters contained identifiably different patterns of student behavior. However, applying these clusters to a test set revealed that cluster interpretations did not generalize well to held out data. The results of subsequent linear regression models suggested that if clustering could be reliably linked to student features, the approach could potentially be used to help drive personalization within the ASSISTments platform.

Limitations of this work include being bound by the hierarchical nature of the data, assumptions inherent to K-Means analysis, and the potential for artificial inflation of model accuracy due to regression to the mean. As it stands, clustering does not necessarily fail as a method of personalization. Understanding the features that are important to each cluster, as well as the overall accuracy of ensembled cluster models and how such accuracy differs with varying values of K, could help to guide the design of learning interventions specific to particular students. However, the reliability of the approach may be extremely sensitive to the quantity and quality of available data, making clustering a difficult approach for personalized learning.

## 6. ACKNOWLEDGMENTS

Thanks to NSF (1440753, 1252297, 1109483, 1316736 & 1031398), U.S. D.O.E. (IES R305A120125 & R305C100024 and GAANN), ONR, and the Gates Foundation.

## 7. REFERENCES

[1] Amershi, S. & Conati, C. 2007. Unsupervised and supervised machine learning in user modeling for intelligent learning environments. Proc 12th Int Conf on Int UI. ACM, 72-81.

[2] Bouchet, F., Harley, J.M., Trevors, G.J., & Azevedo, R. 2013. Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning. *JEDM*. 5(1): 104-146.

[3] Corbett, A.T. & Anderson, J.R. 1995. Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*. 4: 253-278.

[4] Deng, L. & Yu, D. 2014. Deep Learning: Methods and Applications. *Found and Trends in Sig Proc*. 7(3-4): 1–199.

[5] Heffernan, N. & Heffernan, C. 2014. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *Int. J AIED*. 24(4): 470-497.

[6] Pavlik, P.I., Cen, H., & Koedinger, K.R. 2009. Performance Factors Analysis: A New Alternative to Knowledge Tracing. AIED. 531-538.

[7] Trivedi S., Pardos Z.A., & Heffernan N.T. 2011. Clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions. Proc 15th Int Conf on AIED. 377-384.

[8] Zakrzewska, D. 2008. Using Clustering Technique for Students' Grouping in Intelligent E-Learning Systems. In A. Holzinger (Ed.): USAB 2008, LNCS 5298, 403–410.