Paper ID #25702

Using a Data Science Pipeline for Course Data: A Case Study Analyzing Heterogeneous Student Data in Two Flipped Classes

Asuman Cagla Acun Sener, University of Louisville

Asuman Cagla Acun Sener holds B.S. and M.S. degrees in Computer Science and Computer Engineering. She is currently pursuing a doctoral degree in Computer Science at Knowledge Discovery & Web Mining Lab, Department of Computer Science and Computer Engineering, University of Louisville. She is also working as a graduate assistant. Her research interests are educational data mining, visualization, predictive modeling and recommender systems.

Dr. Jeffrey Lloyd Hieb, University of Louisville

Jeffrey L. Hieb is an Associate Professor in the Department of Engineering Fundamentals at the University of Louisville. He graduated from Furman University in 1992 with degrees in Computer Science and Philosophy. After 10 years working in industry, he returned to school, completing his Ph.D. in Computer Science Engineering at the University of Louisville's Speed School of Engineering in 2008. Since completing his degree, he has been teaching engineering mathematics courses and continuing his dissertation research in cyber security for industrial control systems. In his teaching, Dr. Hieb focuses on innovative and effective use of tablets, digital ink, and other technology and is currently investigating the use of the flipped classroom model and collaborative learning. His research in cyber security for industrial control systems is focused on high assurance field devices using microkernel architectures.

Prof. Olfa Nasraoui , University of Louisville

Olfa Nasraoui is Professor of Computer Engineering and Computer Science, Endowed Chair of e-commerce, and the founding director of the Knowledge Discovery and Web Mining Lab at the University of Louisville. She received her Ph.D. in Computer Engineering and Computer Science from the University of Missouri-Columbia in 1999. From 2000 to 2004, she was an Assistant Professor at the University of Memphis. Her research activities include Data Mining/ Machine Learning, Web Mining, Information Retrieval and Personalization, in particular in problems involving large multiple domain, high dimensional data, such as text, transactions, and social network data. She is the recipient of the National Science Foundation CAREER Award, and the winner of two Best Paper Awards, a Best Paper Award in theoretical developments in computational intelligence at the Artificial Neural Networks In Engineering conference (ANNIE 2001) and a Best Paper Award at the Knowledge Discovery and Information Retrieval conference in Seville, Spain (KDIR 2018). She has more than 200 refereed publications, including over 47 journal papers and book chapters and 12 edited volumes. Her research has been funded notably by NSF and NASA. Between 2004 and 2008, she has co-organized the yearly WebKDD workshops on User Profiling and Web Usage Mining at the ACM KDD conference. She has served on the program committee member, track chair, or senior program committee of several Data mining, Big Data, and Artificial Intelligence conferences, including ACM KDD, WWW, RecSys, IEEE Big Data, ICDM, SDM, AAAI, etc. In summer 2015, she served as Technical Mentor/Project Lead at the Data Science for Social Good Fellowship, in the Center for Data Science and Public Policy at the University of Chicago. She is a member of ACM, ACM SigKDD, senior member of IEEE and IEEE-WIE. She is also on the leadership team of the Kentucky Girls STEM collaborative network.

USING A DATA SCIENCE PIPELINE FOR COURSE DATA: A CASE STUDY ANALYZING HETEROGENEOUS STUDENT DATA IN TWO FLIPPED CLASSES

Abstract

This study presents a data science methodology to integrate and explore disparate student data from an engineering-mathematics course. Our methodology is based on exploratory data mining and visualization for analyzing and visualizing raw student data from multiple data sources. The exploratory analysis serves two purposes, 1) it supports the instructor's desire to gain insights into the implementation of a flipped classroom and 2) it serves as a case study for a proposed data science pipeline for educational data. As part of the flipped class, the instructor had students completed assignments in an online homework system before each class meeting and work readiness assessment tests (RATs) at the beginning of each class. The RAT scores were recorded in excel and combined with student performance data that was exported from the online homework system. Paper exams were administered at the end of each unit, and combined with RAT scores, lesson assignment scores, and demographic data. A combination of data mining and classical statistical techniques were used to reveal the trends and peculiarities in the data, without having a specific question or topic to investigate. The data science pipeline which we present has four major stages: *data preprocessing*, *exploratory factor analysis*, *visualization,* and *feature engineering*. Our study revealed some trends and clusters within and across course units. Analysis results show the differences and similarities within the course units and help track learner behavior. A few differences related to gender were found, but prior experience in a course taught using the flipped classroom model did not show a significant difference. Exploratory factor analysis identified two factors from the whole data: *class activities and exam*s (factor 1) and *homeworks and lesson assignments* (factor 2). The discovered factors were found to cluster in two groups within the course units: *Unit 1 to 7* and *Unit 8 to 13*, which has a dividing point at the withdraw date. Results also showed that female students had more class activity scores (i.e. they attended and participated in more classes) than male students. Future work will include collecting more data and generating hypotheses that can be tested using collected data.

1. Introduction

Traditionally, student data in an individual class consisted of homework assignments scores, exam and quiz scores, and project/lab scores. Those scores were usually manually entered in a gradebook. With course materials and assignments moving online and new online educational technology tools being released with great frequency, the landscape of data associated with an individual student in a specific class is changing rapidly; the amount of data recorded for each student in a class is expanding much more rapidly than faculty's ability to assimilate that data. Computed scores and averages that align with the traditional class measures are one way that this data is assimilated; however, there might be other information "hidden" in all that data. This need to discover hidden information has motivated the use of data mining methods in the education domain.

Based on the meta-analysis research paper [1], the authors found that the most popular techniques for educational data mining (EDM) were: clustering, followed by classification, sequential pattern mining, prediction, and association rule analysis. Also, Baker [2] divides EDM research in the following general categories: prediction, clustering, relationship mining, discovery with models, and distillation of data for human judgment. Specifically, for flipped classroom data analysis, several efforts have been reported [3]-[4]. They are mainly focused on comparing student scores of flipped classroom and traditional class methods for the same department and same course, and they are mostly engineering departments [5]. Some studies have performed data analysis on student feedback [6]-[9]. Technology enhanced learning has also led to an explosion in the amount of data collected from student interactions with online learning systems, and this has motivated the use of data and web mining techniques for creating adaptive websites that can recommend content personalized to the student [10]-[11], and that can enable personalized the search for educational content [12]. In other studies [13], researchers analyzed the learner behavior groups by using clustering techniques for MOOC.

Although education data mining has been used successfully in the past, as described above, this paper explores the use of data science strategies to provide insights and understanding to one instructor's implementation of two flipped class rather than to compare the outcomes of a flipped class to a traditionally taught class. The approach explored here can be applied to other courses, regardless of their instructional strategy, though it is most appealing when there is a significant amount of student data collected as part of the class.

When classes include the collection and recording of lots of student data, these data points often exist in disconnected systems (including learning management systems, flat files on instructors' computers and any array of online educational technology systems such as Learning Catalytics, iClickers, Kahot, or Gradescope). The resulting data presents two challenges, namely being *unstructured* and *heterogeneous*, challenges that motivate developing a data science pipeline to streamline and facilitate analysis. *Data science* is the science of data, with an overall goal to explain through available data, processes and objects in a way that is objective and appropriate to making predictions and guiding decisions. A *data science pipeline* is a process that consists of the sequence of steps needed to answer questions, draw insights, or build a decision-making tool from data as part of a data science goal. The steps may include obtaining, cleaning, visualizing, modeling, and interpreting data within a particular domain.

1.1 Objectives

The objective of this study is to integrate and analyze disparate student data (demographics, traditional exam scores, and online homework scores - MyMathLab) to uncover potentially hidden information that can help understand student learning behavior and performance in an active learning environment that uses a flipped classroom instruction model. Specifically, in this analysis we explored the factors of gender and previous participation in a flipped class on student performance.

2. Background

This study presents the results from using data science methodologies to explore all the individual student data collected in two different courses taken by engineering students at a large research 1 university in the midwest. Both courses used the flipped classroom model of instruction.

2.1 The flipped classroom

"*Flipped learning is a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning environment where the educator guides students as they apply concepts and engage creatively in the subject matter.*"
-- Flipped Learning Network (http://flippedlearning.org/definition-of-flipped-learning/)

Descriptions of the flipped classroom model of instruction are many and varied [14]-[17]. Bishop and Verleger [17], define the flipped classroom as an educational technique that consists of two parts: interactive group learning activities inside the classroom, and direct computer-based individual instruction outside the classroom. Talbert [16] explains the flipped classroom by comparing it to the more traditional three phase model of instructional design for a given unit of content: 1) select the concepts (and write learning objectives), 2) use class time to present information on the main topics and concepts, and 3) have students work outside of class on activities intended to bring about mastery of the content (learning objectives). The flipped classroom is an alternative approach where phase 2 and 3 are reversed or flipped such that instructors: 1) select the concepts (and write learning objectives), 2) give students structured out-of-class assignments where they initially encounter the concepts and topics on their own, and 3) have students work on activities during class that assess their basic knowledge and facilitate assimilating the concepts and topics by constructing their own knowledge of the topic.

2.2 Courses and class format

This case study includes data from two math courses, calculus II (course 1) calculus III (course 2), which were taught at large research university in the Midwest. The enrollment in these courses was exclusively engineering students. Both courses were four credit hours and met five days a week. The instructor has some previous experience with flipped classes, but this was the first time the instructor taught these classes using a flipped classroom strategy. The format for each course was as follows:

- Each course was broken down into units or modules, 13 for course 1 and 6 for course 2
- Each unit consisted of multiple lessons. Each lesson was a combination of video materials, textbook reading assignments, and online practice problems designed to introduce unit topics to the students. Lessons were built and assigned in online multimedia homework system (MyMathLab) and a lesson assignment was due before each class meeting (accept for exam days). Each lesson assignment score was exported from the online system to become part of the data set. We use U1L2 to indicate unit 1 lesson 2

- At the beginning of each class meeting students worked a readiness assessment test (RAT) based on the lesson assignment for that class meeting. RATs were on paper and were collected as soon as time expired. They were scored using a rubric that gave 5 points for completely correct, 4 for some correct work, and 3 for some work, even if it was incorrect. Students who were not in class that day received a score of zero. RATs scores were stored in an excel file on the instructor’s computer. In the analysis this is recorded as a class activity score (CA), such that U1L2_CA is the class activity score for unit 1 lesson 2.
- After collecting RATs the instructor provided a mini-lecture, organized around the RAT questions.
- The remaining class period was spent working problems in a team setting, with the instructor having short discussions with the teams and sharing any misconceptions or misunderstandings with the class.
- Before the unit exam students were assigned to complete a unit homework set. This was a set of problems that complemented the problems worked in class. (HWU1 for unit 1 homework).

2.3 Exploratory Factor Analysis

Exploratory factor analysis (EFA) is a statistical method used to uncover the underlying structure of a relatively large set of variables [18]. Traditionally, factor analysis has been used to explore the possible underlying structure of a set of interrelated variables without imposing any preconceived structure on the outcome [19]. To determine the number of factors, Cattell [20] introduced scree plots, which are visual tools used to help determine the number of important components or factors in multivariate settings, such as principal component analysis and factor analysis. The scree plot is examined for a natural break between the large eigenvalues and the remaining small eigenvalues. After applying EFA, factor loadings need to be rotated to become interpretable [21]. There are two main factor rotation methods; orthogonal rotation and oblique rotation. An orthogonal rotation assumes that the factors are uncorrelated, while an oblique rotation assumes that factors are correlated [22]. Oblique rotation was employed in this study since the theory naturally permits inter-correlation between the constructs.

2.4 Visual Data Analysis

An emerging field, blending statistics, data mining, and visualization, visual data analysis comprises a set of methods used for discovering and understanding patterns in large datasets by way of visual interpretation [1]. Bar charts, histograms, scatter plots, social network graphs, stream graphs, tree maps, heat maps, and correlation plots are different techniques that can be used for data visualization [23].

3. Methodology

3.1 Data

Data was collected through the online learning system (MyMathLab), from scores stored in an Excel file on the instructor's computer and the university's institutional data (demographic features of sex and ACT math score). Scores for course 1 and course 2 include:

- scores for each assigned lesson (U1L2 for unit 1 lesson 2),
- homework scores (U1HWK for unit 1 homework),
- RATs (recorded as a class activity score for the class meeting following each assigned online lesson, U1L2_CA for unit 1 lesson 2 class activity score).

The course 1 dataset has scores for 97 students, altogether 335 scores of combined homework, class activities, and exams for each of 13 units. The course 2 dataset has scores for 100 students, altogether 297 scores of combined homework and lesson assignments, class activities, and exams for each of 6 units. This study was approved by the school Institutional Review Board.

3.2 Data Science Pipeline

Figure 1 depicts the data science pipeline with the data flow and stages of our methodology, consisting of four major stages, described below.

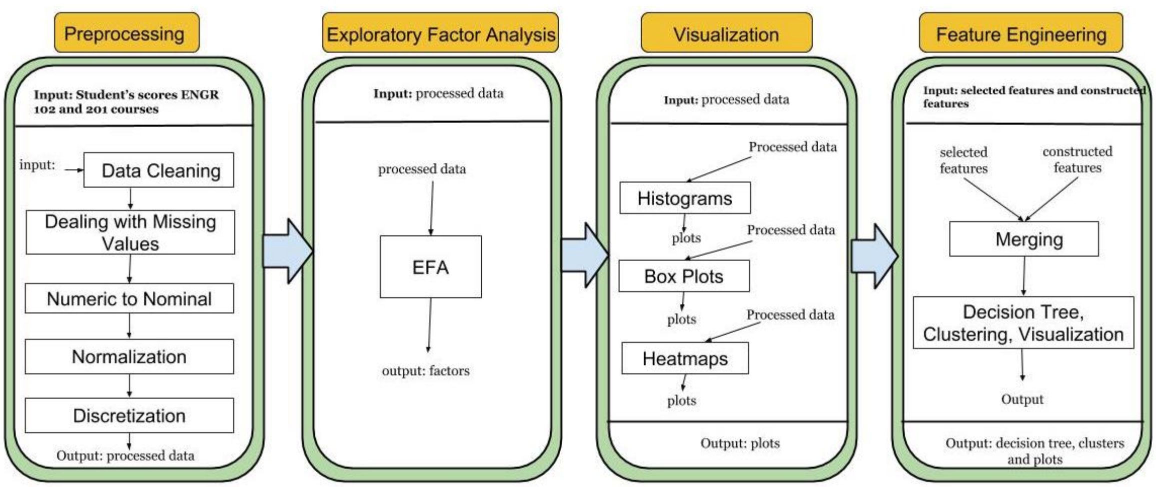

Figure 1: Stages of the proposed data science pipeline.

3.2.1 Preprocessing

Before we analyzed the data, we performed the following data preprocessing steps:

- Cleaning the data by selecting the attributes that are pertinent for our analysis objectives and checking for inconsistencies and missing values.
- Dealing with missing values by replacing a small number of N/A values with zeros; in these cases, the N/A score indicated that the student did not complete and or submit the assignment, for example, the student didn't take an exam.

- Data transformation by: (1) transforming numerical values to nominal attributes for the purpose of building decision trees that are interpretable; (2) normalizing values by centering score data around a zero mean (the mean for the entire class for one activity or exam).

Normalized values allow the comparison between different scores in terms of how they are changing relative to each activity's class average, and are useful for interpreting visualization in a consistent manner across different activity scores; (3) Discretizing score values into the two discrete categories of below and above the mean to obtain interpretable prediction results.

3.2.2 Exploratory Factor Analysis (EFA)

We applied EFA using R libraries readxl, ggplot2, psych, corrplot, and GPArotation.oblique. The fit of the maximum likelihood factor analysis was conducted using the root mean square error of approximation (RMSEA) fit measure and Tucker Lewis Index (TLI). RMESA values less than 0.01, 0.05, and 0.08 indicate excellent, good, and moderate fit, respectively [24]; while TLI values greater than 0.90 indicate good fit [25]. We relied on Tabachnick and Fidell's [26] recommended cut-offs for factor loading values - 0.32 (*poor*), 0.45 (*fair*), 0.55 (*good*), 0.63 (*very good*), 0.71 (*excellent*) - to eliminate poor factor loadings.

3.2.3. Visualization

Histograms, box plots, and heat maps were created using Python libraries, pandas, matplotlib pyplot, ggplot, plotly, numpy, scipy stats, and seaborn. For this study, we created 50 different graphs, only a few are included in this paper.

3.2.4. Feature Engineering

Feature engineering is used when building predictive models where we clearly have an outcome to predict (a discrete class label or continuous variable). Feature engineering can also help in unsupervised learning and preliminaries exploratory analysis to allow us to reveal data stories that may be hidden within the data such as whether there are distinct groups, trends, or correlations. It can also help us build more meaningful visualizations. After exploratory factor analysis and visual data analysis, we constructed new features that are relevant to the questions that emerged in the initial explorations (visualization). Those features included the average of the scores for each student in certain categories, such as exams, and the percentage of class absences of each student (using the class activity scores). The engineered features were used to cluster and build decision tree predictive models for the final score. We used Agglomerative Clustering algorithm for clustering, which recursively merges the pair of clusters that minimally increases a given linkage distance [27]. We used Decision Tree Regressor method for prediction which is a non-parametric supervised learning method used for classification and regression [28]. All methods that are used is an implementation of scikit learn library in Python.

4. Results

4.1. Student Demographics

The dataset included scores for 190 students including 43 females and 147 males. The number of males was almost three times higher than the number of females, 77% of male and 23% of female. Almost half of the students' ACT scores were not available. For the known values, scores were distributed in the range {24-30} with 14% of the students having the highest score range {33-36}. Male students have 28.5 average ACT math score while females had an average of 27.1.

4.2. Exploratory Factor Analysis (EFA)

The RMSEA for the EFA model for course 1 is 0.06 with 90% confidence intervals, and the TLI is 0.91, both indicating a good fit. The RMSEA for the EFA model for course 2 is 0.06 with 90% confidence intervals, and the TLI is 0.88, both indicating a good fit.

Table 1. Exploratory Factor Analysis results of Course 1

	Homework		**Class Activity**		**Exam**
Factor 1	Unit 1 to 3	**Factor 1**	Unit 1 to 7	**Factor 1**	Unit 8 to 13
Factor 2	Unit 4 to 7	**Factor 2**	Unit 8 to 13	**Factor 2**	Unit 1 to 7
Factor 3	Unit 8 to 13				

Table 2. Exploratory Factor Analysis results of Course 2

	Combination of All Scores
Factor 1	Class Activities and Exams
Factor 2	Homework and Lesson Assignments

Table 1 summarizes the results of EFA for Course 1, for three separate EFA for each of the three activity categories' score data subsets, separately. Class Activity Factor 1 includes the last 6 units which range between Units 8 and 13, while factor 2 includes the first 7 unit activities. The factor loadings show that the most significant scores are in factor 1 and they are the review class activities of each unit. Table 2 shows the EFA results for the combined scores for Course 2, showing that Exam scores are grouped with Class Activities and Lesson Assignment scores are grouped with Homeworks. We can conclude that student in-class and out-class scores are in different factors.

4.3. Visualization

Figure 2 shows a visualization of all the collected course data for course 1. The last column of figure 2 is final exam score, the data is sorted by the final exam score. Female students attend lessons more regularly than male students. Continuous streaks of orange to red cells at the end of a student's activity stream likely indicates withdrawal from the course (officially or unofficially), which appears to occur often for students whose early class activity scores and lesson assignment scores are low (dark orange to red). This visualization also allows quickly tracking the status of students as they progress with the course, with a darkening of the light green color, indicating an improvement for the student, and vice versa. In addition to other plots, this visualization shows that male students tend to fail the class more than females. 1 in 7.6 female students drop the class; on the other hand, 1 in 5.5 male students drop this class. Overall drop out ratio is 1 in 5.58 students and 16% of the whole class. Also, with the exception of students who end up dropping the class, there appears to be an improvement trend in class activity scores, lesson assignment scores, homework scores, and exam scores; the right side of the plot shows greener and less orange cells than the left side.

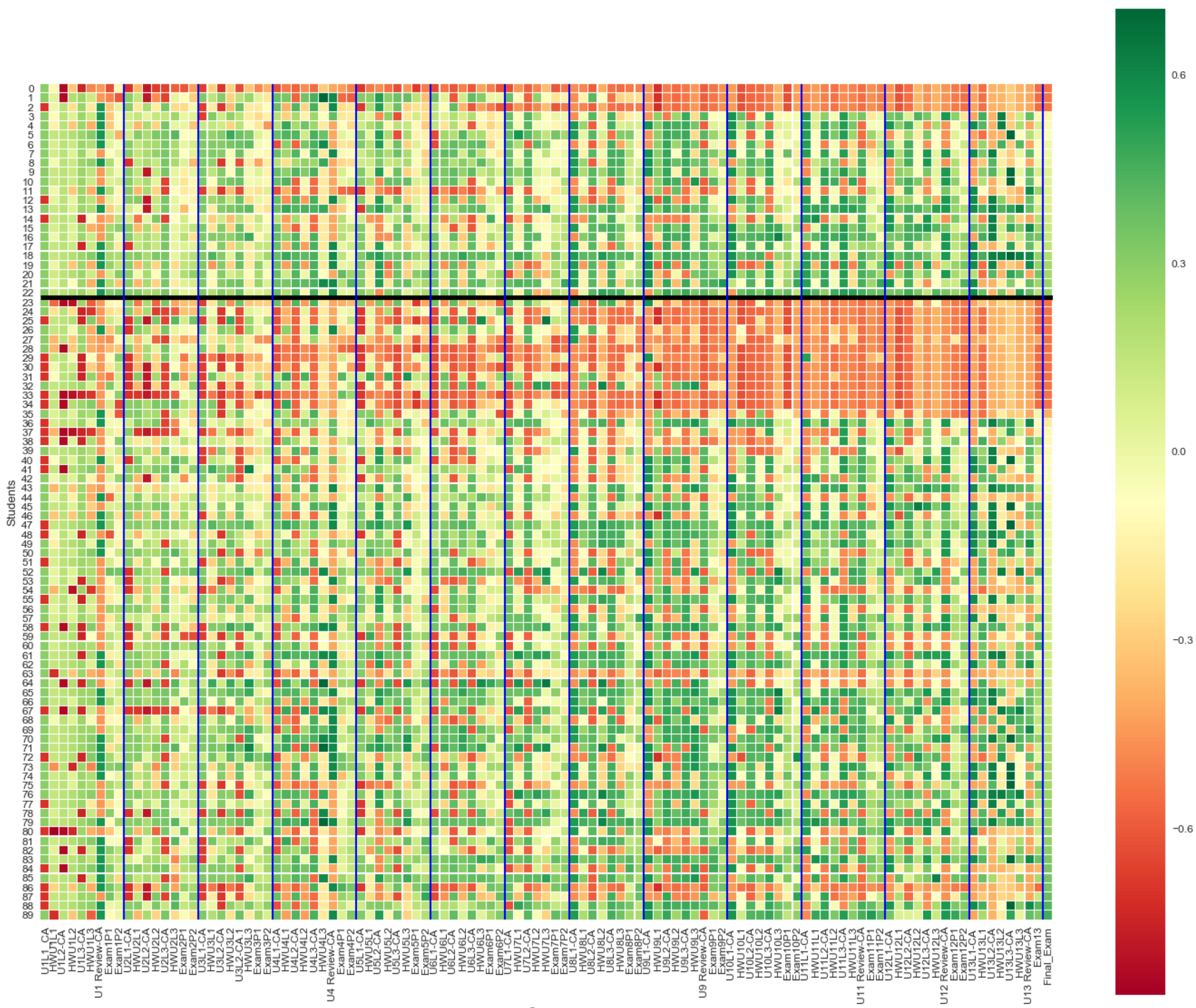

Figure 2: **All Scores Combined for Course 1** ((in columns) vs students (in rows), **sorted by the final exam score within each gender group**. Data is centered to a zero mean. U X is for Unit X; L Y is Lesson Y; HW is homework; CA is class activity and Exam X is for Unit X Exam. Vertical blue lines separate each unit. The horizontal black line divides students in two groups by gender (the female students are in the top part). The thick vertical blue lines separate each unit.

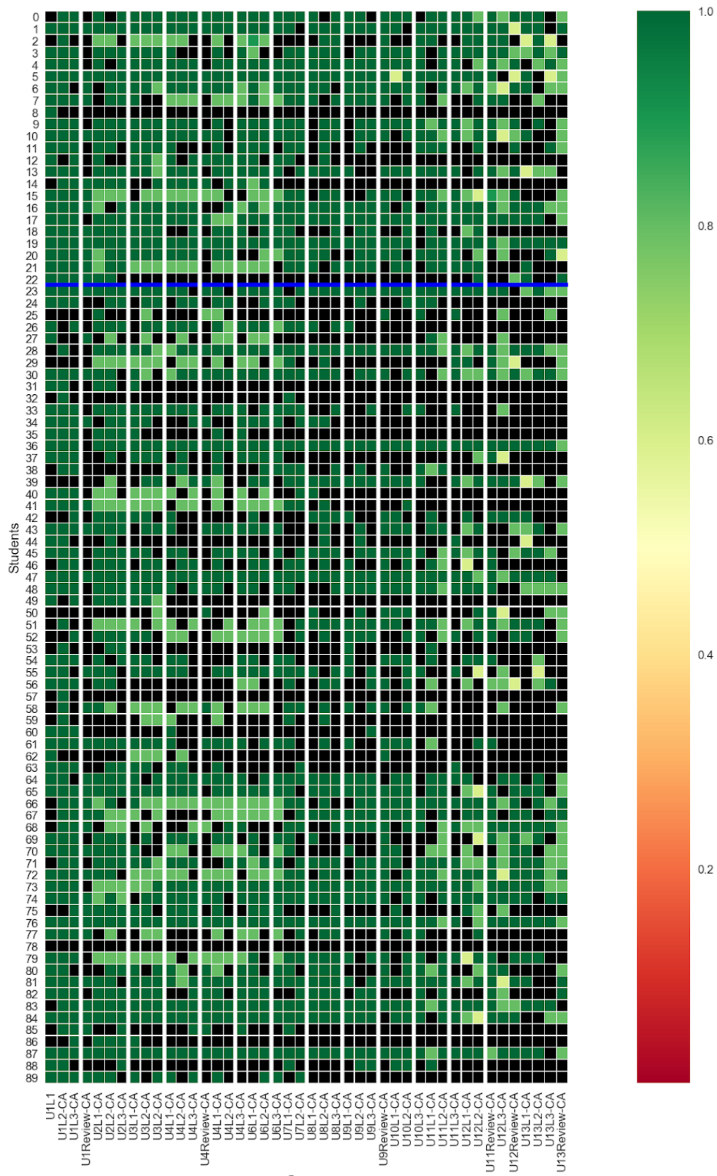

Figure 3: **Class Activity Scores for Course 1** ((in columns) vs students (in rows). Data is not centered to a zero mean. The horizontal blue line divides students by gender (the female students are in the top part). The thick vertical white lines separate each unit. The x axis labels are as follows: U X is for Unit X; L Y is Lesson Y; CA is class activity. Scores range from 0 to 1. Black indicates absence. Female students attend lessons more regularly than male students. Continuous streak of black cells at the end of a student's activity stream likely indicates withdrawal from or dropping the course, which appears to occur often for students who start with more absences early in the course. This visualization also allows quickly tracking the status of students as they progress with the course, with a darkening of the green color, indicating an improvement for the student, and vice versa. It is interesting to turn these visualizations into interactive dashboards to allow a quick tracking of student progress. All of our conclusions from this visualization should later be confirmed using statistical tests.

In the next figure, Figure 4, a visualization of all the course data for course 2 is shown a similar format as figure 2 The visualization is divided by blue horizontal line, where the portion of the blue line is female, and the portion below is mail. Each of these sections is further divided by a horizontal purple line, where below the line is students who took course 1 and therefore have prior experience with flipped classes, and above the line is students who were not enrolled in course 1.

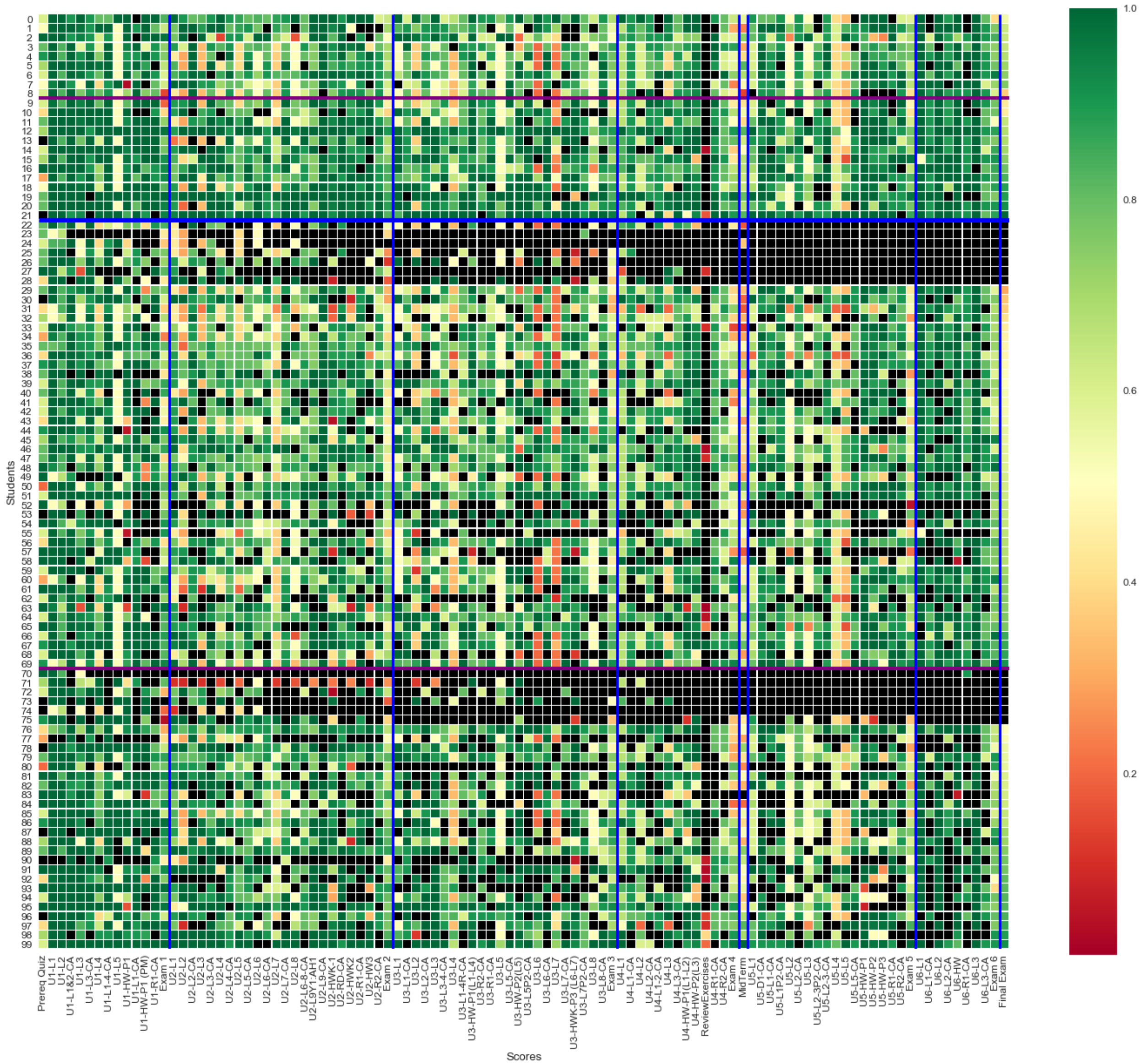

Figure 4: Heat map of **all scores for Course 2** ((in columns) vs students (in rows), **sorted by final exam score within each (gender, prior flipped class) group**. Data is normalized in range (0,1), with 0 being coded as black. Ux is for Unit x; Ly is Lesson y; HW/K is homework.

One interesting observation about this visualization is that you can see students, male and female, with prior flipped experience appear to perform better at in the first several weeks of the course on lesson assignments and class activities. That difference appears to fade further into the semester.

4.4. Feature Engineering

We constructed new features to enrich our data exploration and to predict the final exam score using decision tree models. Constructed features include statistical aggregates of Class Activity scores, Exam scores, homework scores, and absentee percentage. Using these new features, a decision tree analysis was performed. The result of which is shown in figure 5.

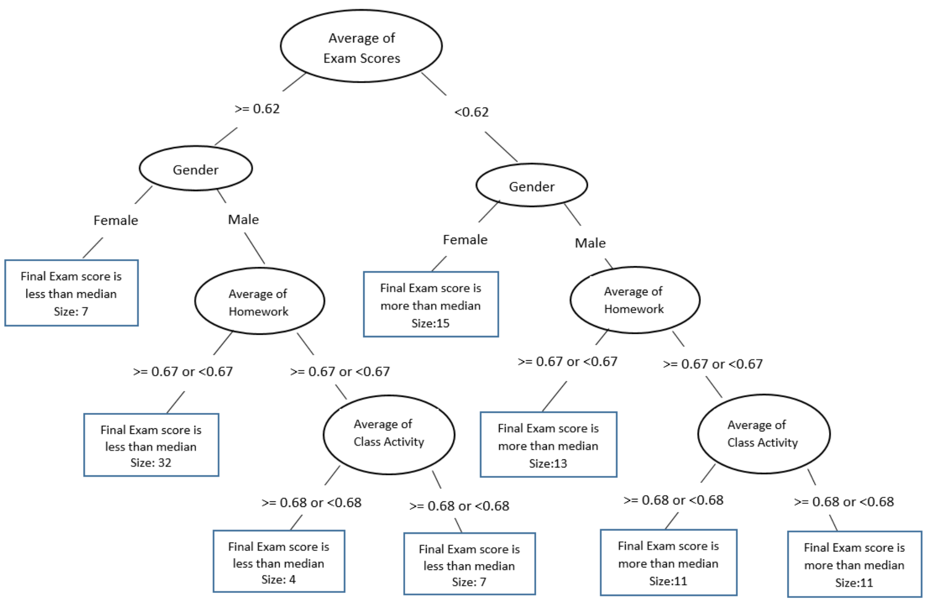

Figure 5: Decision tree using the constructed features for Course 2 for the prediction of the final score (discretized based on the average score as threshold, into low and high ranges). All scores have also been discretized in a similar way. This is a pruned decision tree, built with the J48 algorithm. It has 77% accuracy in predicting the final score. We observe that the gender feature has the highest impact on our prediction model. The second important feature is homework scores and the third one is class activity. Note that this tree is shown as proof of concept of the methodology.

5. Discussion and Conclusion

In this paper, we presented a data science pipeline to analyze the education data set consisting of scores in lessons, homework, exams, etc., in a flipped classroom model for Engineering Students. We used a combination of classical statistical methods with computational visualization and machine learning. Some of the visualizations revealed trends in the increase of scores within and across units, as well as differences based on gender and having taken the flipped classroom before. It could be interesting to turn these visualizations into interactive dashboards to allow a quick tracking of student progress.

To confirm some of our findings about gender and the flipped class factor, we applied the chi-square test of independence. For gender; the p-value is 0.004068 which is less than 0.05; the average score is thus dependent on the gender of students. For the flipped class factor; p-value is 0.6659 which is more than 0.05, the average score is independent of whether student took the flipped class before or not.

We emphasize that visualizations tend to be interpreted subjectively, while rigorous statistical tests remain the best way to verify certain conclusions. On the other hand, visualizations, especially on large data sets, can reveal certain patterns that we may not anticipate, and thus help generate hypotheses to be tested in a later stage using classical methods. While our objective was not to predict the final score, we did build machine learning models to predict this score based on a variety of constructed features. The main goal of these models was to explore which features had the biggest impact on the final score, generally considered as a measure of overall student success in a class. Future work involves improving and constructing new visualizations, as well as continuing some of the hypothesis generation and rigorous statistical testing and modeling. Other approaches such as sequential pattern mining are also needed to support some of the visual inspection of the heat maps. Other data can also be captured to support investigations that leverage data science, based on some of the conclusions we made and unanswered questions.

Our work is a demonstration of the proof of concept that data science, particularly visualization offers a simplified way to interact with big education data sets. For instance, our heat map visualizations, which integrated data from several sources for all students, allow quickly tracking the status of students as they progress within a course, and this status can be interpreted by watching for simple visual clues such as moving from red to green or the darkening of a color, indicating an improvement for the student, and vice versa. It is interesting, as future work, to turn these visualizations into interactive dashboards to allow a quick tracking of student progress. Other extensions include automatic alerts that are integrated with the visualizations.

References:

[1] Z. T. Siti Khadijah Mohamada, "Educational data mining: A review," in The 9th International Conference on Cognitive Science, Malaysia, 2013.

[2] R. S. Baker, "Data Mining for Education," in International Encyclopedia of Education (3rd edition), Oxford, 2012.

[3] C. G. Merrett, "Using Textbook Readings, YouTube Videos, and Case Studies for Flipped Classroom Instruction of Engineering Design," in Proc. 2015 Canadian Engineering Education Association (CEEA15) Conf., Canada, 2015.

[4] D. N. A. G. M. S. Kenneth A. Connor, "Faculty Development and Patterns of Student Grouping in Flipped Classrooms Enabled by Personal Instrumentation," in 2017 ASEE Annual Conference & Exposition, Columbus, 2017.

[5] H. G. S. Gabriela C. Weaver, "Design, Implementation, and Evaluation of a Flipped Format General Chemistry Course," Chemical Education Research, vol. 9, no. 92, pp. 1437-1448, 2015.

[6] A.-M. K., M. O.-L. Jonathan Verrett, "I flipped my tutorials: a case study of implementing active learning strategies in engineering," in Proc. 2015 Canadian Engineering Education Association (CEEA15) Conf., Canada, 2015.

[7] F. A. Fei Geng, "Biotechnology labs reinvented through experiential learning: Enhancing student outcomes through the “flipped lab”," in 2015: Proceedings of the Canadian Engineering Education Association (CEEA) Conference, Canada, 2015.

[8] C. L. H. L. A. R. Mary Lou Maher, "Flipped Classroom Strategies for CS Education," in SIGCSE '15 Proceedings of the 46th ACM Technical Symposium on Computer Science Education, Kansas City, 2015.

[9] Sahin A., Cavlazoglu, B., & Zeytuncu, Y. E. (2015). Flipping a College Calculus Course: A Case Study. Educational Technology & Society, 18 (3), 142–152.

[10] Khribi, M. K., Jemni, M., & Nasraoui, O. (2008, July). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on (pp. 241-245). IEEE.

[11] Khribi, M. K., Jemni, M., & Nasraoui, O. (2015). Recommendation systems for personalized technology-enhanced learning. In Ubiquitous learning environments and technologies (pp. 159-180). Springer, Berlin, Heidelberg.

[12] Zhuhadar, L., & Nasraoui, O. (2008, November). Semantic information retrieval for personalized e-learning. In Tools with Artificial Intelligence, 2008. ICTAI'08. 20th IEEE International Conference on (Vol. 1, pp. 364-368). IEEE.

[13] Kerrie A. Douglas, Peter Bermel, Md Monzurul Alam, Krishna Madhavan (2016). Big data characterization of learner behaviour in a highly technical MOOC engineering course. Journal of Learning Analytics, 3(3), 170–192.

[14] Jensen, J. L., Kummer, T. A. & Godoy, P. D. d M. Improvements from a flipped classroom may simply be the fruits of active learning. CBE-Life Sci. Educ. 14**,** ar5 (2015).

[15] DeLozier, S. J. & Rhodes, M. G. Flipped Classrooms: a Review of Key Ideas and Recommendations for Practice. Educ. Psychol. Rev. (2016). doi:10.1007/s10648-015-9356-9

[16] Talbert, R. Flipped Calculus A Gateway to Lifelong Learning in Mathematics. (Routledge, 2015).

[17] Bishop, J. L. & Verleger, M. A. The flipped classroom: A survey of the research. in ASEE National Conference Proceedings, Atlanta, GA 30**,** (2013).

[18] Norris, Megan; Lecavalier, Luc (17 July 2009). "Evaluating the Use of Exploratory Factor Analysis in Developmental Disability Psychological Research". Journal of Autism and Developmental Disorders. 40 (1): 8–20.

[19] D. Child, The essentials of factor analysis, second edition, London: Cassel Educational Limited, 1990.

[20] C. R. B, "The scree test for the number of factors," in Multivariate Behav. Res. , Urbana-Champaign, University of Illinois, Urbana-Champaign, ILl, 1966, pp. 1:245-76.

[21] J. W. Osborne, "What is Rotating in Exploratory Factor Analysis?, "Practical Assessment, Research & Evaluation, vol. 20, no. 2, p. 7, 2015.

[22] J. D. Brown, "Choosing the Right Type of Rotation in PCA and EFA," Shiken: JALT Testing & Evaluation SIG Newsletter, vol. 13, no. 3, pp. 20-25, November 2009.

[23] Michael Friendly (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization"

[24] MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods. 1996;1(2):130-149.

[25] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model A Multidiscip J. 1999;6(1):1-55.

[26] Tabachnick BG, Fidell LS, others. Using multivariate statistics. 2001.

[27] Zhang, et al. "Graph degree linkage: Agglomerative clustering on a directed graph." 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012

[28] L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees”, Wadsworth, Belmont, CA, 1984