Online Education Performance Prediction via Time-related Features

Bowei Hong, Zhiqiang Wei, Yongquan Yang*
Department of Computer Science and Technology,
Ocean University of China
Qingdao, China
e-mail: i@hongbowei.com, weizhiqiang@ouc.edu.cn, i@yangyongquan.com

***Abstract*—In this work, we studied time management behavior, performance of students, and their association in online learning. We propose three novel time-related features, i.e., video-watching frequency, video-watching interval and learning efficiency. Leveraging these features, we train classifiers for identifying study habits of students and assessing the impact of behavior on performance. Experimental results on dataset from our online learning platform demonstrate that the proposed features are effective for learning behavior description. Comprehensive experiments show that our method can get promising performance.**

Keywords—online learning; data mining; performance prediction; time-related feature

I. Introduction

Over the past decades, increasing online education platforms have made knowledge become more and more easy to be accessed[1], [2]. Most platforms track interactions of students among video lectures, assessments, and social networking[3], which has motivated a number of recent studies focused on understanding how online learning users learn (e.g., [4]–[7]). Also, researchers have drawn attention to the relationship between those interactions which has far-reaching implications to methods for improving low completion rates[7], such as personalized content delivery[3] and instructor analytics[8].

Our work is motivated by these studies. In our research, we suppose to meet two objectives:

- O1. Identifying study habits of learners such as prefer watching lectures at certain time or intend to take long breaks before starting with a new lecture.
- O2. Assessing the impact of behavior on performance, i.e. those patterns identified in O1.

In study O1 and O2, we identify three features that are important to capture: one based on the duration of the time user finished one video lecture and started a new one, one based on video-watching time distribution over hours, and one based on the duration that a user used to finish all video lectures of one course. In particular, we first extract these features from cleaned datasets. Leveraging these features, we train classifiers for O1 and O2. We find that a series of study habits are indicative of initiative on material of students, and are significantly associated with their performance in the course. As another finding, we identify study habits that are consistent with slow-paced video watching, and reveal that these are discriminatory in terms of poor performance in courses. These findings show that time management behavior information is useful in some situations, especially when other type extraction of the clickstream events sometimes does not reveal as much information as they usually do.

Compared with other work (related work section), we make two contributions. First, we present novel features for describing learning behavior, which is useful in identifying study habits and for performance prediction. Second, we predict user learning performance without clickstream sequences and assignments grades used in previous research works[4], [9], [10] which is limited when online learning platforms require users to perform a standard watching process, such as no playback and skip, and associate these fundamental behaviors with student performance.

II. Related Work

Various works have investigated in online education, student video-watching analysis and performance prediction.

Online education studies. With the proliferation of online education in recent years, there have been a number of analytical studies on these platforms. Some have focused on a more general analysis of all learning modes, e.g., [11], [12] studied learner engagement variation over time and across courses. Others have focused on specific modes. In terms of forums, [6] analyzed the decline in participation over 73 courses. In terms of assessments, [13] focused on predicting assignment completion in MOOC. Our work is fundamentally different from these in that is explores the association between behavior with two modes: videos and assessments.

Video-watching analysis. Much existing work studying learner video-watching behavior [5], [9], [14]has focused a variety of user interaction characteristics(e.g., re-watching session, playback, pause, fraction spent) with few time-distribution-related information. [15] also studies video-watching clickstreams, except video-watching logs they used data from social networking (e.g., play and pause logs on the video, as well as post and comment threads on the forums). In this way, they are focused on a more complete view of the learner experience, including transitions both within and across these different learning modes. Also, the short work in [16] presents the possibility of representing video clickstream

Supported by the CERNET Innovation Project under Grand No.NGII20150201; the Shandong Province key research and development plan under Grand No.2016ZDJS09A01; the Shandong Province Science development plan under Grand No.2014GGX101005.

978-1-5090-5507-4/17/$31.00 ©2017 IEEE
ICIS 2017, May 24-26, 2017, Wuhan, China

trajectories as a time series of playback position over real (UNIX) time for a single video used in a FLIP course, which opens the possibility of applying time series similarity methods (e.g., dynamic time warping) to further compare watching behavior.

Performance prediction. Researchers have developed predictors for performance of students in the context of traditional education settings. Collaborative filtering (CoF) algorithms[17], probabilistic graphical models (PGMs) such as Hidden Markov Models (HMMs) and Bayesian networks[18], [19], decision tree classifiers[20], and factorization machines (FM)[21] have been applied for prediction in this setting, in which work requires information beyond learning behaviors (e.g., course difficulty, age range). References [9], [22] have used learning behavioral data for performance prediction. However, they mainly rely on continually clickstream events with video players, once online education platform controls the interaction for specific reasons, information revealed via clickstream events would be limited, which might cause prediction result accuracy decline. Compared with these works, ours is unique in that (i) it focuses instead on relating time-distribution behavioral data- video-watching behavior - to performance, and (ii) it focuses on use only time-related data on performance prediction, which is click-event-free.

III. METHOD

We proposed novel time-domain features. All three features are extracted from video-watching logs, consisting of video-watching frequency, video-watching interval and learning efficiency. The system flow of the processing framework is described in Fig. 1. The database and corresponding ground truth labels are prepared following instruction introduced in experiment section. Each sample in the database extracted from four datasets: course data, student enrollment data, video-watching logs, grades records, which are combined together in the preprocess part.

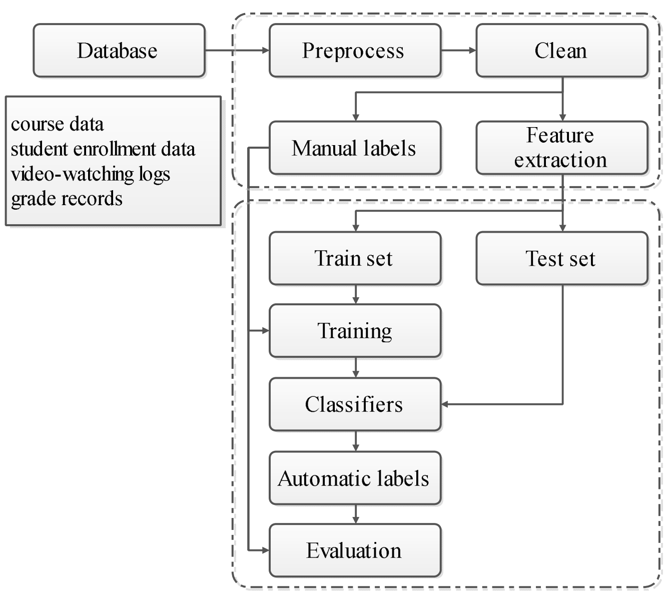

Fig. 1. Flowchart of Our System

Preprocess and clean. After getting data from online learning platform, we turned to process these data. Fig. 2 illustrates the detail workflow of data preprocess and data clean. And we carried out the following work in data preprocess:

Since each student has a unique identifier and a series of matching video-watching logs in the database, we collected all the user-video pairs into groups which refer to courses on the online platform. From those pairs that are grouped together, we were able to fully reconstruct the watching trajectory for each user-course pair.

In terms of data clean, we followed these procedures:

We remove all user-course pairs which did not have a grade record. Then we dropped the pairs which have a grade record but have an uncompleted user-course trajectory. Most of these situations were caused by errors happened during system migration or upgrade.

Also, those records contain at least one "null" parameter entry have been removed from our datasets. From the remainder, we discounted all entries that have got time sequence conflicts, e.g., a user- course pair that has user-video pairs of two different lectures in the same time records, since different lectures belonging to a same course are required to watch in order. For last step, we identified outliers from our dataset via boxplot tool, and removed them lest they affected the offset of overall features.

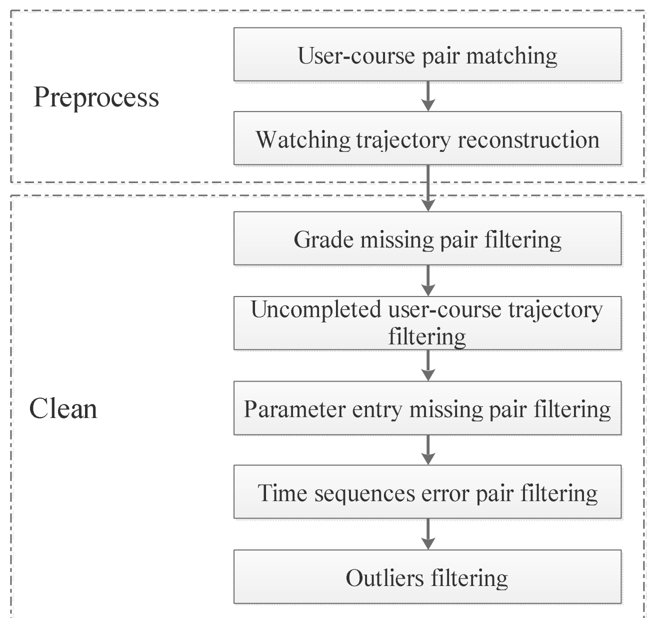

Fig. 2. Workflow of Data Preprocess and Clean

Feature extraction and normalization. From video-watching logs, we obtained three types of data: video-watching frequency, video-watching time interval, and learning efficiency, which have logged time management behaviors of users during the course. The details of feature description were characterized in Table I.

TABLE I. DETAIL DESCRIPTION OF LEARNING BEHAVIOR FEATURE EXTRACTED METHODS, THE COLUMN OF "LENGTH" REPRESENTS FEATURE DIMENSIONS

No	Feature Dimensions		
	Feature Name	***Description***	***Length***
1	Video-watching frequency $F_i, i \in [1,24]$	The frequency of Video-watching behavior occurs in i^{th} hour during whole course learning.	24
2	Video-watching interval $W_i, i \in [1,8]$	The (real) time the user spent during the i^{th} break between the i^{th} lecture and the $i+1^{th}$ lecture.	8
3	Learning efficiency *diff*	the duration a user spent on finishing all video lectures of the course	1

After applying Min-Max Normalization by (1), we extracted learning behavior feature from the dataset after data preprocessing and data clean.

$$z_i = \frac{\max(x) - x_i}{\max(x) - \min(x)} \tag{1}$$

Where $x = (x_1, ..., x_n)$ and z_i is i^{th} normalized sample.

Training and evaluation. After features extraction and normalization, our dataset is split into training set and test set. Training set along with corresponding manual labels is applied to train classifiers, and then test set is used to obtain automatic labels with trained classifiers. Assessment categories used as the outputs of classification are not unified. In this paper, we separate students into two groups: do well / poorly according to rules given in [10]. With the ground truth, we do the evaluation.

In order to validate the new features we proposed, we trained three classifiers based on three standard classifier algorithms, which are support vector machine[23], decision tree[24] and random forest[25] respectively. Above all algorithms, we applied their default parameters.

IV. EXPERIMENT AND ANALYSIS

A. Dataset and Evaluation Metrics

Using 70% of the available data, we trained classifiers, and the remaining 30% of the data was used to test the performance of them. Our dataset comes from one of the courses we have instructed on an online training platform. In order to guarantee that the information provided by learning records is sufficient enough, we have chosen a higher enrollment course –Project Cost Management - as our data source over multiple offerings. After applying data preprocess procedures described in method section, records of 339 users have been removed from our dataset, which means we have got a dataset contains online-learning logs of 834 users eventually collected from January 2014 to September 2014. All user data has been privacy cleaned.

To obtain the ground truth for our dataset, we programed to classify students into two categories according to their course grades. Because grades are often used to summarize how well a student was able to understand and apply the knowledge conveyed in a course, it is proper to adopt grades as our ground truth. Table II shows the number of people per performance category.

TABLE II. NUMBER OF PEOPLE PER PERFORMANCE CATEGORY

	People per Performance Category	
	Good Performance	***Good Performance***
Sample Number	559	275
Percentage of Sample	67.03	30.97

Moreover, we used three primary metrics in reporting the performance of the classifiers: sensitivity, specificity, and area under the ROC because of the unbalanced sample size for each category (See [26] for a discussion). In addition, ROC curves are drawn for visualization purpose. Our metrics may be interpreted as follow:

Sensitivity – proportion of good performances that are correctly classified

Specificity – proportion of poor performances that are correctly classified

AUC – A single scalar value representing the overall performance of the classifier.

B. Result and Analysis

Table III lists the classification performance of three classifiers using our features. All three classifiers get decent performance. Among which Random Forest get the highest sensitivity but the lowest specificity, that makes its AUC value ranks second. Meanwhile, SVM improves specificity 31% than Random Forest, but worsens sensitivity by 2%. Moreover, it is interesting that, compare to Random Forest, AUC of C50 descents to 0.78 while making specificity 14% improved with only a 5% decrease of specificity. In particular, we note that SVM appears to work quite well achieving AUC of 0.85 and specificity of 0.56.

TABLE III. PERFORMANCE OF THREE CLASSIFIERS

No		Classifiers		
	Performance	***SVM***	***C50***	***Random Forest***
1	***Sensitivity***	0.93	0.90	0.95
2	***Specificity***	0.56	0.39	0.78
3	***AUC***	0.85	0.78	0.81

Fig. 3 illustrates the performance comparison on our dataset, in which three algorithms with probabilistic outputs are presented in ROC curves. In real application, the performance with FPR between 25% and 50% attracts the most attention, and SVM tends to achieve the largest TPR in this range, which corresponds with its achievement of AUC.

We now analyze how different features can help predicting performance of students by training each feature respectively. Since SVM performs well in term of specificity and AUC, we choose it to implement the analysis. Table 4 shows that all the features have contributions to performance prediction.

Furthermore, video-watching frequency contributes most in term of AUC and specificity achieving 0.73 and 0.31 respectively, while video-watching interval and learning efficiency are most useful in term of sensitivity. However, this result not only indicates that all features have contributions to performance prediction, but also reveals that the way students manage study time affects their course performance most in term of time related factors.

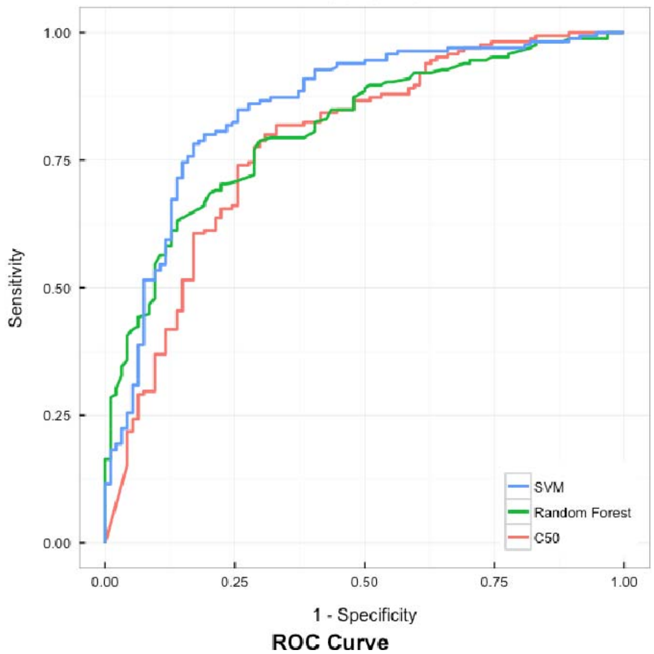

Fig. 3. Performance Comparison via ROC Curves

TABLE IV. FEATURE CONTRIBUTION ANALYSIS

No		Feature		
	Performance	***Sensitivity***	***Specificity***	***AUC***
1	***Video-watching frequency***	0.93	0.31	0.73
2	***Video-watching interval***	0.98	0.00	0.60
3	***Learning efficiency***	0.98	0.03	0.50

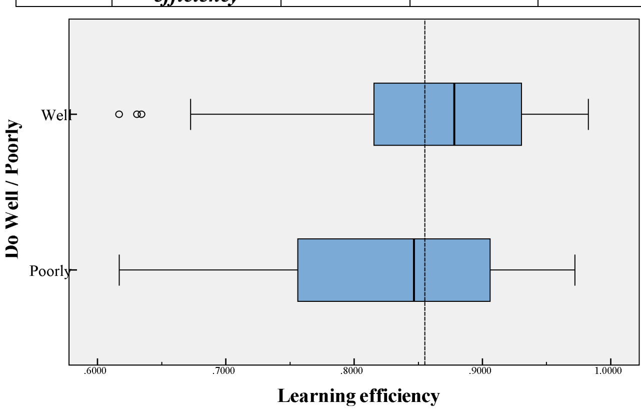

Fig. 4. Boxplot of Learning Efficiency over Users Where User Group with Good Performance Gets Lower Variances.

In Fig. 4, we plot how the learning efficiency varies across students. The top shows group "Do poorly", the mean is 0.830 with SD = 0.098. At the bottom we plot the learning efficiency across users "Do well", and the SD drops substantially (0.082), with the mean higher slightly (0.863). The dotted line indicates the mean (0.855) of entire users in terms of learning efficiency. According to formula (1), the values note that, users who finish lectures with less time than average tend to do well in final examination.

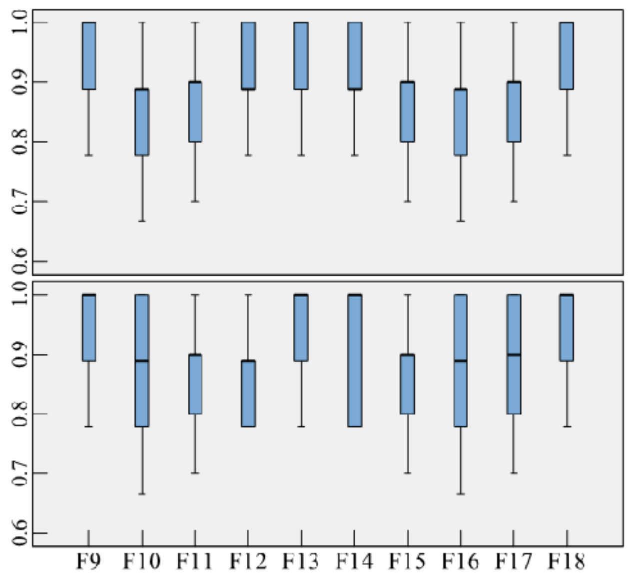

Fig. 5. Boxplot of Video-watching Frequency

In Fig. 5, we plot video-watching time distribution over users from 9th hour to 18th hour of a day with which period the most user study. It shows that users do well more likely to study at 10th, 11th, 15th, 16th, and 17th hour of a day. Moreover, they spend their time more regularly than the other group does, which means a lower variance.

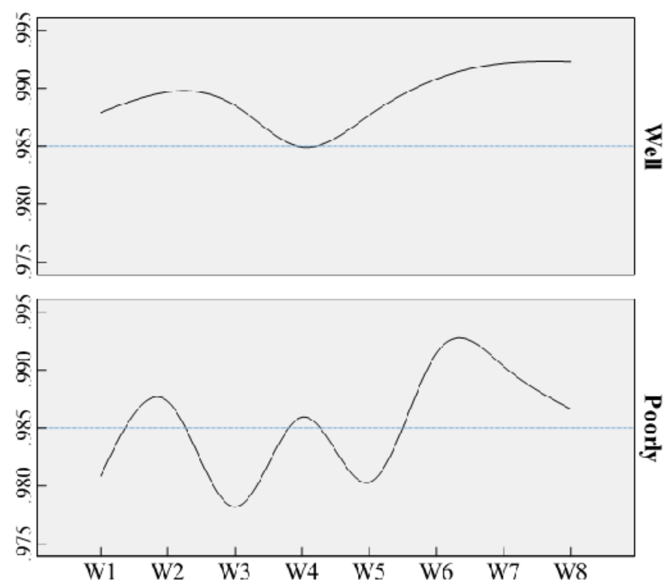

Fig. 6. Depictions of Average Intervals over Users

In Fig. 6, we plot the average intervals over users. The blue dotted lines show the interval means of all video-watching interval windows over all users, and black lines represent the interval means over well and poorly performance group respectively. It is clear that the top black line totally over the dotted line, while the bottom black line fluctuates around it, which indicates that allocates time between two lectures more evenly tends to do better in course.

V. Conclusions

In this work, we studied time management behavior, performance of students, and their association in online learning. We proposed three novel time-related features: Video-watching frequency, Video-watching interval and Learning efficiency. With dataset from our online learning platform, we accomplished two goals: (1) we present novel features for describing learning behavior, which is useful in identifying study habits and for performance prediction; (2) we accomplished the prediction of the final grades of students exclusively based on their video-watching-time-related behaviors. Since our prediction considers time-related features exclusively, this benefit suggests that our method is useful in situations where the interactions with learning lectures are limited, e.g., platform requires no skip and playback at the first time user watches video.

Acknowledgment

We acknowledge Dr. Lei Huang, Dr. Zhen Li and Dr. Hao Liu for their instructive suggestions and valuable advice.

References

[1] T. Väljataga, H. Põldoja, and M. Laanpere, "Open online courses: Responding to design challenges," *Stanford Univ. H-STAR Institute, USA; to Assoc. Profr. Jukka M. Laitam{ä}ki, from New York Univ. USA, to Profr. Yngve Troye Nord. from Lillehammer Univ.*, p. 68, 2011.

[2] C. Tekin and M. van der Schaar, "eTutor: Online Learning for Personalized Education," *arXiv Prepr. arXiv1410.3617*, 2014.

[3] C. G. Brinton, R. Rill, S. Ha, M. Chiang, R. Smith, and W. Ju, "Individualization for Education at Scale: MIIC Design and Preliminary Evaluation," *Learn. Technol. IEEE Trans.*, vol. 8, no. 1, pp. 136–148, 2015.

[4] T. Sinha, "' Your click decides your fate ': Leveraging clickstream patterns from MOOC videos to infer students ' information processing & attrition behavior Bachelor of Technology degree in Computer Science and Engineering School of Computing Science and Engineerin," 2014.

[5] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller, "Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos," *Learn. Scale, 2014*, pp. 31–40, 2014.

[6] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong, "Learning about social learning in MOOCs: From statistical analysis to generative model," *IEEE Trans. Learn. Technol.*, vol. 7, no. 4, pp. 2013–2014, Dec. 2013.

[7] C. G. Brinton and M. Chiang, "Social learning networks: A brief survey," *2014 48th Annu. Conf. Inf. Sci. Syst.*, pp. 1–6, 2014.

[8] K. Stephens-martinez, M. A. Hearst, and A. Fox, "Monitoring MOOCs : Which Information Sources Do Instructors Value ?," 2014.

[9] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining MOOC Clickstreams: On the Relationship Between Learner Behavior and Performance," *arXiv Prepr. arXiv1503.06489*, 2015.

[10] Y. Meier, J. Xu, O. Atan, and M. van der Schaar, "Predicting Grades," pp. 1–13, 2015.

[11] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," *WWW '14 Proc. 23rd Int. Conf. World wide web*, pp. 687–698, 2014.

[12] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing Disengagement : Analyzing Learner Subpopulations in Massive Open Online Courses," *Lak '13*, p. 10, 2013.

[13] J. Cheng, C. Kulkarni, and S. Klemmer, "Tools for predicting drop-off in large online classes," *CSCW '13 Proc. 2013 Conf. Comput. Support. Coop. Work Companion*, pp. 121–124, 2013.

[14] C. G. Brinton and M. Chiang, "MOOC Performance Prediction via Clickstream Data and Social Learning Networks," *IEEE Conf. Comput. Commun.*, 2015.

[15] T. Sinha, N. Li, P. Jermann, and P. Dillenbourg, "Capturing 'attrition intensifying' structural traits from didactic interaction sequences of MOOC learners," *EMNLP 2014 Work. Model. Large Scale Soc. Interact. Massively Open Online Courses*, pp. 42–49, 2014.

[16] J. M. Aiken, S. Lin, S. S. Douglas, E. F. Greco, D. Brian, M. D. Caballero, and M. F. Schatz, "Student Use of a Single Lecture Video in a Flipped Introductory Mechanics Course," *Phys. Educ. Res. Conf. Proc. 2014*, 2014.

[17] A. Toscher, "Collaborative filtering applied to educational data mining," *KDD Cup*, no. January, 2010.

[18] Z. Pardos and N. Heffernan, "Using HMMs and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset," *... Mach. Learn. Res. W ...*, no. 2010, 2010.

[19] M. U. F. K. O. Z. A. a P. P. T. D. J. Sheikh M.B., "A Bayesian approach to online performance modeling for database appliances using gaussian models," *Proc. 8th ACM Int. Conf. Auton. Comput. ICAC 2011 Co-located Work.*, pp. 121–130, 2011.

[20] N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," in *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual*, 2007, p. T2G--7.

[21] M. Sweeney, J. Lester, and H. Rangwala, "Next-term student grade prediction," in *Big Data (Big Data), 2015 IEEE International Conference on*, 2015, pp. 970–975.

[22] T. Sinha, P. Jermann, N. Li, and P. Dillenbourg, "Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions," *CoRR*, vol. abs/1407.7, 2014.

[23] C.-C. Chang and C.-J. Lin, “{LIBSVM}: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27:1--27:27, 2011.

[24] M. Kuhn, S. Weston, N. Coulter, M. Culp, and R. Quinlan, “C50: C5.0 Decision Trees and Rule-Based Models.” 2015.

[25] A. Liaw and M. Wiener, “Classification and Regression by randomForest,” *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[26] F. M. Harper, D. Moy, and J. A. Konstan, “Facts or Friends?: Distinguishing Informational and Conversational Questions in Social Q&A Sites,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 759–768.