SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students

Rui Wang, Gabriella Harari†
Peilin Hao, Xia Zhou, and Andrew T. Campbell
Dartmouth College, The University of Texas at Austin†,
{ruiwang, hao, xia, campbell}@cs.dartmouth.edu,
gabriella.harari@utexas.edu

ABSTRACT
Many cognitive, behavioral, and environmental factors impact student learning during college. The *SmartGPA study* uses passive sensing data and self-reports from students' smartphones to understand individual behavioral differences between high and low performers during a single 10-week term. We propose new methods for better understanding study (e.g., study duration) and social (e.g., partying) behavior of a group of undergraduates. We show that there are a number of important behavioral factors automatically inferred from smartphones that significantly correlate with term and cumulative GPA, including time series analysis of activity, conversational interaction, mobility, class attendance, studying, and partying. We propose a simple model based on linear regression with lasso regularization that can accurately predict cumulative GPA. The predicted GPA strongly correlates with the ground truth from students' transcripts ($r = 0.81$ and $p < 0.001$) and predicts GPA within ± 0.179 of the reported grades. Our results open the way for novel interventions to improve academic performance.

Author Keywords
Smartphone sensing; data analysis; academic performance; behavioral trends

ACM Classification Keywords
H.1.2 User/Machine Systems; I.5 Pattern Recognition

General Terms
Algorithms, Experimentation.

INTRODUCTION

College life is complex. Students have to balance going to classes and performing well academically with competing demands for their time and energy, such as extracurricular activities, busy social lives, working because of financial concerns, being members of under-represented minorities, fitting in on campus (e.g., first-generation college students), dealing with friends and families, and trying to stay active and physically and mentally healthy during the ebb and flow of the term's workload and commitments. As a result, succeeding in a demanding educational environment is challenging.

When we think of "academic performance" we usually associate it with educational outcomes best represented by a student's cumulative GPA. This measure typically captures a continuous assessment of a student's academic achievement in terms of results from assignments, quizzes, tests, midterms, and final examinations as they move through their college years. Academic performance is linked to a student's intellectual curiosity and ability (e.g., as measured by IQ), their drive and motivation, the educational environment, health, prior test results (e.g., SATs), and personality traits (e.g., conscientiousness). There is no general agreement, however, on why students with similar academic capability at the same institution do better or worse than one another. It seems likely that students' everyday behavioral patterns (e.g., study habits, class attendance, time management, sleep patterns, partying behavior) significantly contribute to individual differences in academic performance among students. This raises several questions. Are there distinct differences in the behavioral patterns of high (e.g., GPA ≥ 3.5) and low performers (e.g., GPA ≤ 3) at the same college? If such behavioral differences exist could we use these correlations as a basis for predicting academic performance? How do different psychological characteristics, such as personality, mental health, affect, and stress collectively contribute to GPA?

This paper makes the following contributions. First, we propose new methods to automatically infer *study* (i.e., study duration and focus) and *social* (i.e., partying) *behaviors* using passive sensing from smartphones. Next, we use time series analysis of these and other behavioral states derived from the StudentLife dataset [43], a longitudinal study of college students, to find what behaviors significantly impact term and cumulative GPA. We use this behavioral analysis as input to model the individual differences between high and low performers in a population of undergraduate students at Dartmouth College. Third, in order to understand changes in behavior students experience across the term we propose two new behavioral metrics: 1) *behavioral slope*, which captures the direction of behavioral change (i.e., increases or decreases in class attendance) and magnitude of the behavioral change (e.g., steep or gradual changes in attendance) over the complete term, as well as the first and second half of the

UbiComp '15, September 07-11, 2015, Osaka, Japan

DOI: http://dx.doi.org/10.1145/2750858.2804251

term; and 2) *behavioral breakpoints*, which capture the specific points in the term when a student's behavioral pattern shows a directional change (i.e., the point when their class attendance increases or decreases). The time series analysis of student behavioral streams and these change metrics are used as input to correlation analysis and prediction of GPA. Finally, we propose for the first time a model that can predict a student's cumulative GPA using automatic behavioral sensing data from smartphones. We use the *Lasso* (Least Absolute Shrinkage and Selection Operator) [40] regularized linear regression model as our predictive model. Our prediction model indicates that students with better grades are more conscientious, study more, experience positive moods across the term but register a drop in positive affect after the midterm point, experience lower levels of stress as the term progresses, are less social in terms of conversations during the evening period, and experience change in their conversation duration patterns later in the term. The predicted GPA strongly correlates with the ground truth with $r = 0.81$ and $p < 0.001$, mean absolute error (MAE) of 0.179, and R^2 of 0.559, which measures the goodness of fit of a model and indicates that our model explained 56% of the variance in students' GPAs. We do this without the use of any prior data that has been traditionally used for academic assessment [15,22,25], such as IQ and standardized test results (e.g., SAT scores). As a result, our work opens the door to predict academic performance using passive sensing data from smartphones.

The StudentLife study [43] looked at correlations between academic performance and the averages of the low level sensor data (i.e., activity, conversation, and mobility) for all students across a term. StudentLife did not, however, study the time series of each individual behavior, nor analyze the individual differences distinguishing high and low performers. The SmartGPA study advances the state-of-the-art by inferring new behaviors, proposes new behavioral change metrics, discovering new correlations, and showing for the first time that passive smartphone sensing data can be used to accurately predict GPA.

RELATED WORK

In the computer science community, many efforts have been made to predict grades from students' self-report data and e-learning behaviors using various machine learning models. However, only a few studies have examined the relationships between students' performance and sensed behaviors. The StudentLife study [9, 43] found correlations between students' GPAs and automatic sensing data obtained from smartphones. In addition, Watanabe and colleagues [44, 45] investigated the correlations between scholastic performance and face-to-face interaction among students during break times using a wearable sensor device. Our research extends this work by building a predictive model of academic performance based on students' self-reports and sensed behavior features obtained from their smartphones.

In the fields of education and psychology, much research has focused on identifying the predictors of college student's academic performance. Overall, the existing studies tend to focus on whether students' personality traits (e.g., extraversion, conscientiousness), lifestyle behaviors (e.g., physical activity, sociability, sleep), and mental states (e.g., stress, positive affect) are related to their course grades or GPA. However, the existing research findings are primarily based on students' self-reports (i.e., one-time surveys asking about general physical activity or sleep tendencies), which may be susceptible to a range of limitations. For example, self-report data have significant drawbacks, such as being disruptive, time consuming, being subject to recall biases, memory limitations, and socially desirable responding (for a review see [26]). Thus, one aim of our study is to use unobtrusive and longitudinal measures of students' lifestyle behaviors to predict performance. Next, we review the existing research that links academic performance with students' personality, behaviors, and emotions.

Personality. Research that examines the links between academic performance and personality tends to adopt the Big Five personality framework [20], which consists of five broad traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness. Taken together, a meta-analytic review of the literature on academic performance and personality suggests that student performance is associated with agreeableness, conscientiousness, and openness to experience [28]. However, some studies also find extraversion and neuroticism to be negatively associated with performance [16]. Some researchers [8] suggest that personality traits may differentially impact academic performance; for example, by impairing performance in the case of neuroticism, or increasing academic achievement in the case of conscientiousness.

Physical Activity. The majority of research focused on academic performance and physical activity tends to suggest that grade averages are higher among students meeting health guidelines for moderate-vigorous physical activity [42]. However, a large cross-sectional study of health behaviors of students in forty U.S. colleges and universities found that more than half of the students (58%) did not meet public health recommendations for moderate-vigorous physical activity [42]. Another self-report based study [19] has found that physical activity self-reports are not associated with student GPAs. This study, however, focused on health science graduate students whose physical activity reports met or exceeded the recommended levels for adults, suggesting that the results may not generalize to other student populations. Only a few studies that we know of have found relationships between academic performance and sensor-based physical activity measures. One study measured student physical activity using a sensor armband in addition to self-reports and found that changes in physical activity were associated with GPA [30]. Specifically, total step count was associated with increases in GPA, whereas reported moderate physical activity was associated with decreases in GPA [30]. The StudentLife study measured student physical activity using accelerometer data from students' smartphones and found that lower overall GPAs was associated with greater average levels and variability of activity durations aggregated over a term [43].

Sociability. Behaviors related to sociability (e.g., duration or frequency of engaging in conversation, partying, and spending time alone or with others) have been difficult to study, which has led to differences in the way sociability is operationalized. For instance, a meta-analysis of predictors of college performance found that social involvement (e.g., social integration, involvement in campus activities) was associated with higher GPAs among college students [31]. However, night outings (i.e., social events such as partying, movies) have been associated with poorer performance [18]. In addition, social support has been linked to higher academic performance among college freshmen [13]. The StudentLife study measures sociability using conversation data inferred from the audio collected by smartphone's microphone [43]. It shows that greater conversation durations aggregated over a term was associated with higher spring term GPAs.

Sleep. The majority of research focused on academic performance and sleep patterns tends to suggest that grade averages are higher among students meeting guidelines for good sleep habits [42]. However, a large cross-sectional study of health behaviors of students in forty U.S. colleges and universities found that only a quarter of students (24%) actually met public sleep recommendations [42]. In addition, some studies that examine the relationship between performance and sleep duration find a negative relationship between self-reported number of hours slept and students' GPA [17]. However, other studies have found a quadratic relationship between performance (i.e., cumulative GPA) and total sleep duration, such that too little or too much sleep is associated with poorer performance [39]. Other research studies have found that wake-up and bed times are important for performance, such that later bed and wake-up times are associated with poorer performance [39, 41]. In addition, variability in sleep behaviors (e.g., bed times, wake times, total sleep duration) has been linked to performance, such that greater variability it associated with poorer performance [39].

Class Attendance and Studying. In general, a meta-analytic review of college performance found that academic-related skills (e.g., study skills and habits) were associated with higher GPAs among college students [31]. Research that examines academic behaviors has also found that absenteeism and class attendance predict academic performance, such that students who attend class more often performed better than those who missed class [8, 11, 18]. For example, a meta-analysis of studies that examine the relationship between class attendance and performance found attendance to be strongly related to class grades and GPA among college students [12]. However, the StudentLife study [43] found no correlation between class attendance and academic performance.

Emotions. Relatively few studies have examined the relationship between academic performance and positive affect. Those studies that have focused on affect found that positive affect was associated with higher grades and GPAs, while negative affect during the second half of the semester was associated with lower grades and GPAs [32] . On the contrary, the relationship between academic performance and students' stress has received more attention. In general, college students tend to report more stressful daily hassles [34]. The existing research tends to show that moderate stress is associated with decreases in students' GPAs [30], and that perceived stress during the end of the semester is associated with lower GPAs [29]. Previous research tends to find a curvilinear relationship between stress and performance, such that too little or too much stress is associated with poorer performance [39].

Performance Prediction. Previous research [15] aimed at predicting performance has used a neural network model to predict student's grades from their placement test scores. Various data collected from entering students are used in [25] to predict student academic success using discriminant function analysis. [22] proposes a regression model to predict the student's performance from their demographic information and tutor's records. [33] applies web usage mining in e-learning systems to predict students' grades in the final exam of a course. In [48], the authors propose an approach based on multiple instance learning to predict student's performance in an e-learning environment. Recent work [38] showed that they can predict a student is at risk of getting poor assessment performance using longitudinal data such as previous test performance and course history. To the best of our knowledge there is no work on using passive sensor data from smartphones as a predictor on academic success.

ACADEMIC PERFORMANCE DATASET

In this paper, we use a subset of the StudentLife dataset to analyze and predict academic peformance. The StudentLife dataset is a large, longitudinal dataset that is publicly available [43]. The dataset is collected from 30 undergrads and 18 graduate students over a 10-week term in spring 2013. Each student takes three classes during a term at Dartmouth College. The dataset includes over 53 GB of continuous sensing data from smartphones, including: 1) *objective sensing data*: sleep (bedtime, duration, wake up), face-to-face conservation duration, face-to-face conversation frequency and physical activity (stationary, walk, run); 2) *location-based data*: location, co-location, indoor/outdoor mobility, and distance covered; 3) *other phone data*: light, Bluetooth, audio, Wi-Fi, screen lock/unlock, phone charge, and app usage. The dataset also comprises 32,000 daily self-reports covering affect (PAM [27]), stress, exercise, mood, loneliness, social and study spaces; and pre-post surveys including PHQ9 depression scale [23, 24, 37] , UCLA loneliness scale [35], positive and negative affect schedule (PANAS) [46], perceived stress scale (PSS) [10], big five personality [20], flourishing scale [14], and the Pittsburgh sleep quality index [6]. Finally, the dataset includes academic assessment data, such as, class information, deadlines, academic performance (i.e., grades, term GPA, cumulative GPA), class attendance rates (from phone location data), Piazza usage data [4], and student dinning history including time, location, and cost.

In this paper, we only use undergraduate students' (N=30) data because only undergraduates have GPAs. In contrast,

Table 1: Sensing and EMA data.

sensing data	activity duration audio inferences (voice/noise/silence) conversation frequency conversation duration distance covered indoor mobility sleep duration location
EMAs	stress positive affect

Table 2: Psychological surveys.

personality	openness conscientiousness extraversion agreeableness neuroticism
mental health	PHQ-9 (pre and post) perceived stress scale (pre and post) UCLA loneliness scale (pre and post) flourishing scale (pre and post)

graduate students [5] do not have GPAs and only receive High Pass, Pass, Low Pass or No Credit for their classes.

Table 1 summarized the automatic sensing data and EMA data we use from the broader StudentLife dataset. Automatic sensing data captures daily behaviors. The EMA data captures positive affect and stress level. Table 2 summarized the pre and post psychological surveys data we use.

For details on the StudentLife study and how behavioral states (e.g., sleep, face-to-face conservation) are inferred see [43].

ASSESSING STUDY AND SOCIAL BEHAVIOR

The StudentLife dataset provides a number of low-level behaviors (e.g., physical activity, sleep duration, and sociability based on face-to-face conversational data) but offers no higher level data related to study and social behaviors, which are likely to impact academic performance. In what follows, we discuss how we attribute meanings or semantics to locations – called behavioral spaces as a basis to better understand study and social behaviors. That is, we extract high level behaviors, such as studying (e.g., study duration and focus) and social (e.g., partying) behaviors by fusing multiple sensor streams with behavioral spaces.

Behavioral Spaces

The StudentLife dataset has two types of location data: GPS and Wi-Fi location. Wi-Fi location is determined using Wi-Fi scanning, which indicates what Wi-Fi access points (APs) are around. The APs are then mapped to specific buildings or area of buildings (e.g., libraries) [43]. We hand label each on-campus building with semantically meaningful labels such as study areas (e.g., libraries, specific cafes where students study), Greek houses, classrooms, gyms). These labels provide clues about a student's behavior at these behavioral spaces. Importantly, we associate a number of attributes with behavioral spaces to give them more contextual meanings, specifically: *(i) dwell time*: the amount of time a student spends at these locations – study area, dorm, party or social spaces; *(ii) activity*: the percentage of stationary labels from the activity classifier [43] among all activity inferences when a student is at specific areas; and finally *(iii) audio*: the percentage of silent labels among all audio inferences from the classifier [43]. For example, in study areas or the classrooms, a high level of stationary labels (e.g., not interacting with their phone) might coarsely be associated with being "focused" on studying or paying attention in class and high levels of silence might indicate focused work and vice versa. We use behavioral spaces and their attributes to better assess study and social behavior, as discussed next.

Study Behavior

Each student takes three classes, which are scheduled at specific periods during the week [2]. Class periods fall into three categories: 65-minute periods three times weekly, 50-minute periods four times weekly, and 110-minute periods twice weekly. In addition, each class has an additional X-period of 50 minutes that a lecturer may or may not use. The earliest classes start at 8.45 AM and the latest finishes at 5.50 PM. Students' transcripts indicate what classes they took. The registrar office has the schedule and location for each class. We use location, date (i.e., weekday M-F) and time to automatically determine if a student attends a class or not, checking the dwell time at the location at least equals 90% of the scheduled period (e.g., 110 minutes). Using this approach the phone can automatically determine the classes a student is taking and their attendance rates. Figure 1(b) shows class attendance rates across the term.

We use behavioral space information to determine study behavior. We heuristically determine if a student's dwell time at a study areas (e.g., library, labs, study rooms, cafes where student primarily work) is at least 20 minutes. We consider periods shorter that 20 minutes are less likely to be real study periods. In addition to dwell time, we use activity and audio attributes to determine a student's level of focus at a study area. The value of activity indicates how often the phone moves – the person is either moving around in the study area or stationary but using the phone. We consider a number of scenarios. If a student is in a study (e.g., a library) and moves around we consider this contributes to a lack of focus. If the phone is mostly stationary in a study area, we consider this contributes to focus. We also use the audio attribute to determine the level of ambient noise in study areas. We consider quiet environments may contribute to study focus and noisy environments do not. Figure 1(a) shows the changing study duration and focus for all the students across the term. In term of focus, a higher activity value indicates that the student moves around less and thus is more focused and a higher audio value indicates that the student is in a quieter environment which is more conducive to being focused. We do not combine these values but use them as independent variables in the analysis section. We acknowledge that both activity and audio attributes can only represent coarse estimations of study focus. For example, noisy environments in cafes where students study may suit certain personalities and be more conducive to studying than quiet libraries. We also cannot determine if a student is actually studying or on online social networks. Furthermore, if students study in a group their conversational data would be considered as a noisy environment.

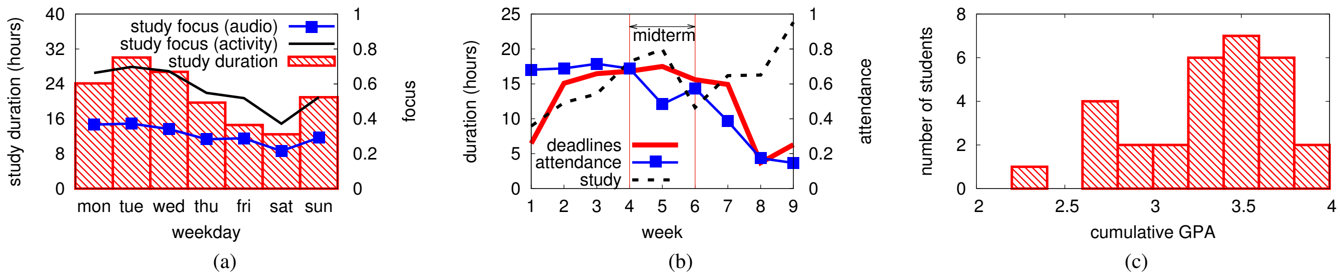

Figure 1: (a) Weekday average total study duration per student and study focus trends. (b) Weekly study duration and lecture duration trends with contrast to the number of deadlines. Note, the number of deadlines is scaled. (c) Cumulative GPA distribution.

Finally, our inference accuracy is based on labeling a large number of well-known study areas at Dartmouth. If students study outside of these areas our data would not reflect that. In addition, we do not have observational data that a student studied while in a labeled study areas – they could be occupied quietly reading a sports blog in a library or napping in cafe where students usually study. Even with these limitations we argue that behavioral spaces combined with dwell times, activity and audio attributes provide a new unobtrusive, if coarse estimation of study duration and focus.

Social behaviors

While the original StudentLife dataset uses the term "social behavior" to mean the number of face-to-face conversations between students, we extend this to include other higher level social behaviors, such as: How often do students party during the week or across the term? How long do they party?

Dartmouth is located in a small college town in Hanover, New Hampshire with few other partying alternatives for drinking other than fraternities and sororities. There are three big drinking nights that many undergraduates attend at Dartmouth College [1]: Wednesday, Friday, and Saturday nights. Wednesday night is when the Greek houses – particularly fraternities – hold their weekly "meetings" (a colloquial term for parties). Both Friday and Saturday are big party nights on campus. Sunday is a day of rest where students buckle down to academics to make progress on assignments before the start of the academic week. Importantly, much of the legal and illegal drinking occurs at these parties, which are located in the basements of the fraternities. Frat parties are open to all students across campus and consist of playing a drinking game called pong with paddles and dance music. There have been a lot of discussions in the press about the safety issues around such a social scene and the new president of the college is wisely trying to create alternative venues for students to party [3].

We consider behavioral spaces (e.g., Greek houses, dorms) and their attributes to infer if a student is partying. If a student is in a party we assume that they will be moving and around acoustic sound of conversation or music. We also consider the day of the week as being significant for the fraternity and sorority parties (i.e., Wednesday, Friday and Saturday). We discard dwell times under 30 minutes at partying locations.

We partition each Greek house dwell periods (i.e., visit or stay) into 10-minute windows and calculate audio and activity attributes. We hypothesize that the audio and the activity attributes should be significantly different when the student is partying or not partying. We use k-means clustering [47] to find the partying thresholds for both the audio (e.g., music or being surrounded by a large group of people) and activity (e.g., dancing) attributes. Figure 2(a), shows that a student is more likely to be in a party when the audio attribute (i.e., the percentage of *silent* labels) is below 40%. Surprisingly, we did not find significant differences in the activity attribute. By fusing audio, dwell time, and location we can distinguish if a student is partying even if they live in the fraternity/sorority. To validate our party inference method, we compare the daily inferred party duration each week with the known party days across the term. Figure 2(b) shows our inferred party data for all the students for each weekday averaged over the 10-week term in terms of the number of hours partied. Clearly Wednesday and Friday are the big nights on campus. In addition, Thursday and Saturday are also popular party nights. Sunday and Monday are not party nights and students are likely catching up with academics. Our data from smartphones strongly aligns with the party weekly pattern ground truth as discussed earlier. Figure 2(c) shows the partying trends across the full term. The party season peaks during the second week of term, steadily drops until after the mid term when it picks up at week 7, when there is a campus-wide spring festival called Green Key Weekend. Finally, dormitories are another places where students socialize. Clearly, these events are not large parties. We use the same approach as discussed above to determine if a student is socializing at a dorm.

CAPTURING BEHAVIORAL CHANGE

In what follows, we discuss behavioral change features extracted from the low-level automatic sensing (e.g., sleep duration) and EMA data (e.g., stress) and high-level study and social behaviors discussed in the previous section. We create time series of each behavior for each student. We use data preprocessing [43] to convert the behavioral data in various forms to a uniform time series format. The behavior time series samples each behavior each day. After the data preprocessing, each time series summarizes a different behavior (e.g., physical activity, conversation frequency and duration, sleep, social behavior, and study behaviors). In order to understand behavior changes across the term, we propose two features: *behavioral slope*, which captures the magnitude of change (e.g., increase or decrease in sleep) over the complete

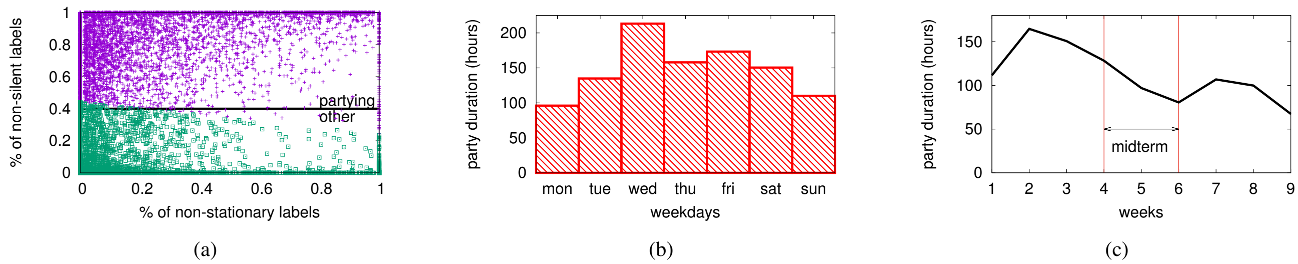

Figure 2: (a) shows clustering of audio and activity feature. 32567 audio and activity pairs are clustered into two clusters. We define the cluster denoted by "+" as the *party* cluster, which contains 9921 pairs. We can clearly see that we can apply simple thresholding on audio attribute ($\geqslant 0.4$) to find the party cluster. (b) shows that Wednesdays, Fridays, and Saturdays are primary party nights, which is in line with the reality [1]. (c) shows students party less during midterms and finals.

term as well as the first and second half of the term for all students – from the start of term to the midterm point, and then from the midterm point to the end of term; and *behavioral breakpoints*, which capture the specific points in the term where individual behavior change occurs – the number of breakpoints a student experiences indicates the rate of change that occurs. In addition to looking at a student's behavior during a day, we partition a day into three epochs, as describe in [43]. Specifically, we label the period between 12am and 9am as the *night epoch*, 9am to 6pm as the *day epoch*, and 6pm to 12 am as the *evening epoch*.

Behavioral Slope

We are interested in quantifying behavioral change of students during the term. For example, is a student more or less active, social, studious, etc., as the term progresses. We capture the behavioral change by computing a slope for each behavioral time series (e.g., indoor mobility, stress, affect) for each student using linear regression. The value of the slope indicates the direction and strength of behavioral changes. A positive slope with a greater absolute value indicates a faster increase in behavioral change (e.g., partying). In contrast, a negative slope with a greater absolute value indicates a faster decreasing behavior level (e.g., class attendance). For example, consider the number of independent conversations a student has each day as a time series over the term. A slope $= 0$ means the student has the same number of conversations each day across a complete term – this is highly unlikely. A slope < 0 means the student has fewer conversations as the term proceeds. And finally, a slope > 0 means that the student has an increasing number of conversations as the term proceeds. The slope of a behavior allows us to take into account the dynamics of behavior and understand individual differences among students. As discussed in the StudentLife study [43] the midterm period (shown in Figure 1(b) as week 4 and 5) is a significant milestone in the term. We select the "midterm point" as a point to measure behavioral slope up to and then after. This point is the center day of the midterm period and mid point of the complete 10 week term period. The workload students experience increases from the beginning of term, as shown in in Figure 1(b). After midterms students have projects and larger assignments culminating in final exams. We partition the behavioral time series at the midterm point and use two linear regressions to fit time series from the beginning of term to the midterm point (i.e., first half of term) and from the midterm point to the end of term (i.e., second half of term). We use the terms *pre-slope* and *post-slope* to capture students' behavioral change during the first and second half of the term, respectively. In addition, we compute a *term-slope* for each behavior taken over the complete term for all students.

Behavioral Breakpoints

The pre-slope and post-slope points are used collectively for all students to understand behavioral change. However, many students may change behaviors at different timescales than the midterm point. Students may enact or experience change (e.g., attend class more or less, study more or less) for many different reasons. For example, some students may change their study behaviors early in the term to adapt to increasing workload, whereas others may react later. We compute "behavioral breakpoints" for each student using the time series of each of their behaviors. We can find a day in the term, before and after which the student's behavior change patterns differ. We call this day a behavioral *breakpoint*. For example, consider a student that spends a similar number of hours studying each day. However, after a certain day the student spends more and more time studying. We consider the point of change as a breakpoint. Many factors influence breakpoints. Different behaviors may have different breakpoints (e.g., an increase or decrease in stress, affect, and studying). The day of the breakpoint may indicate how quickly a student enacts changes because of an event. We use two linear regressions to fit the data and use the Bayesian information criterion (BIC) [36] to select the best model. BIC is a model selection criterion that selects a model with good predictive performance using as few model parameters as possible. Lower BIC value indicates a better model. In our analysis, we consider a good piecewise fitting model as the lowest BIC among all piecewise models and also lower than the single regression model. If the single regression model is selected, the breakpoint is set to the last day. Using per-student behavioral breakpoints we can analyze the rate of changes occurring across the term and understand individual differences.

RESULTS

In this section, we first conduct correlation analysis to find which time series features have significant connection with academic performance, specifically, the spring term and cumulative GPA. Figure 1(c) and Figure 3(c) show the distribution of the cumulative and term GPA for the students in the study. Cumulative GPA indicates a student's overall long term academic performance. The spring term GPA captures how a student performs in a single 10-week term. After correlation analysis we discuss our model for predicting cumulative GPA.

Correlation Analysis

To best understand the relationship between student behaviors, emotions, mental health, and personality, and academic outcomes we conduct Pearson correlation analysis. We identify a number of strong and significant correlations.

Spring Term GPA. The mean of spring term GPA is 3.3306 and the standard deviation is 0.7983. The skewness, however, is -1.7725, meaning that most students receive high GPAs for the term and only a small portion of students get low GPAs (Figure 3(c)). In the StudentLife study [43], we found the spring term GPA negatively correlates with the means of indoor mobility and positively correlates with the conversation frequency and duration. Here we present results from newly designed features, as shown in Table 3.

We find a number of significant correlations between study and social behavior and GPA. In terms of social behavior, we find that students who spend more time partying at fraternities or sororities are less likely to have high GPAs ($r = -0.398, p = 0.029$). In addition, students that socialize more at their dorms rather than fraternities prior to the midterm point are more likely to have higher GPAs ($r = 0.363, p = 0.049$). In terms of study behavior, we find students who spend more time studying have higher GPAs ($r = 0.381, p = 0.038$). In addition, students who show an increase in the amount of time they devote to studying prior to the midterm point ($r = 0.397, p = 0.030$) are more likely to have better grades.

In terms of other behaviors inferred from automatic sensing, we find a decrease in physical activity throughout the term particularly after the midterm point negatively correlates with spring term GPA ($r = -0.576, p = 0.001$), meaning that students who experience a decrease in their physical activity levels are more likely to have higher GPAs. We find similar indoor mobility trends and activity trends during the day, night, and evening epochs. Interestingly, students who have higher GPAs tend to increase their indoor mobility prior to the midterm point ($r = 0.423, p = 0.020$), and decrease their indoor mobility after the midterm point ($r = -0.515, p = 0.004$). In addition, these high achievers tend to spend more time in their dorms throughout the term ($r = 0.437, p = 0.016$). Finally, we find the perceived stress scale negatively correlates with the spring term GPA, meaning students who are less stressed are more likely to have higher GPAs ($r = -0.405, p = 0.050$).

Cumulative GPA. The mean of cumulative GPA is 3.4215 and the standard deviation is 0.3978. The GPA distribution is shown in Figure 1(c). In the StudentLife study [43], we found the cumulative GPA negatively correlates with the means of activity duration and indoor mobility; and positively correlates with the number of Bluetooth co-locations. Here we present results from newly designed features, as shown in Table 4.

Table 3: Spring Term GPA Correlations.

	features	**r**	**p-value**
automatic sensing	activity term-slope	-0.551	0.002
	activity post-slope	-0.576	0.001
	activity night term-slope	-0.431	0.017
	activity night post-slope	-0.654	< 0.001
	activity day term-slope	-0.411	0.024
	activity day post-slope	-0.442	0.016
	activity evening term-slope	-0.485	0.007
	conversation freq night breakpoint	0.379	0.039
	indoor mobility term-slope	-0.606	< 0.001
	indoor mobility pre-slope	0.423	0.020
	indoor mobility post-slope	-0.515	0.004
	indoor mobility night term-slope	-0.529	0.003
	indoor mobility night pre-slope	0.365	0.047
	indoor mobility night post-slope	-0.543	0.002
	indoor mobility day term-slope	-0.568	0.001
	indoor mobility day post-slope	-0.371	0.048
	indoor mobility evening term-slope	-0.552	0.002
	dorm duration term-slope	0.437	0.016
	social duration dorm pre-slope	0.363	0.049
	party duration mean	-0.398	0.029
	study duration mean	0.381	0.038
	study duration pre-slope	0.397	0.030
survey	perceived stress scale (post)	-0.405	0.050

We find a number of significant correlations between study and social behavior and GPA. Students who spend more time studying are more likely to have higher GPAs ($r = 0.518, p = 0.003$). In addition, students who are more focused in terms of their activity (i.e., their phone is more stationary, $r = 0.430, p = 0.018$) and audio (i.e., study in quieter environments, $r = 0.380, p = 0.038$) attributes are more likely to have higher GPAs. The study focus trends, however, show that students who have higher GPAs tend to have a decreasing focus (i.e., activity attribute) before the midterm point and prefer to study at locations that are not quiet (e.g., cafe area). The attendance rate does not correlate with the GPA as discussed in the StudentLife study [43]. However, we find that a change of attendance before midterm positively correlates with the GPA, meaning that students who increase their attendance before midterm point are more likely to have higher GPAs ($r = 0.470, p = 0.009$).

In terms of other behaviors inferred from automatic sensing, we find significant correlations between GPA and behavior change trends. Students whose physical activity level increases more before midterm ($r = 0.418, p = 0.022$) or decreases after midterm ($r = -0.449, p = 0.015$) are more likely to have higher GPAs. Similarly, students who move around indoors more toward the midterm ($r = 0.425, p = 0.019$) or move around more after the midterm ($r = -0.426, p = 0.021$) are more likely to have higher GPAs. Looking at the overall behavioral changes, students whose indoor mobility increases more slowly or decreases throughout the term are more likely to have higher GPAs ($r = -0.387, p = 0.035$). In term of daily conversation duration, students who have increasing conversation durations after the midterm point are more likely to have higher GPAs ($r = 0.443, p = 0.016$). Regarding the conversation frequency, students with later breakpoints (i.e., changed their daily conversation frequency pattern later in the term) are more likely to have higher GPAs ($r = 0.641, p < 0.001$ for

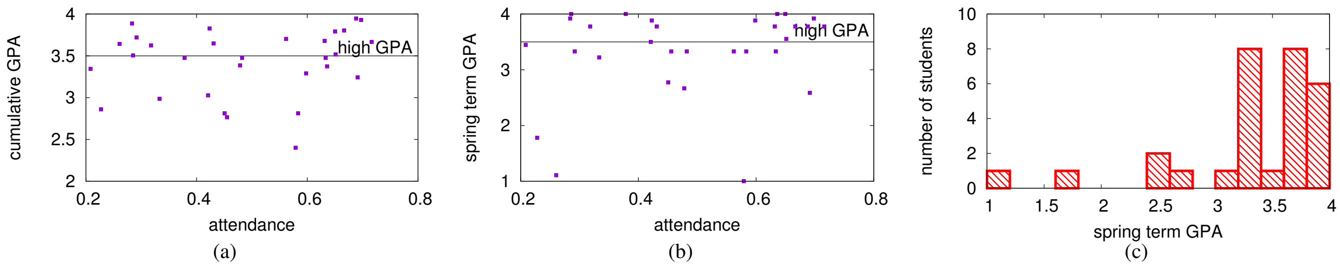

Figure 3: (a) Attendance and the cumulative GPA. (b) Attendance and the spring term GPA. (c) Spring term GPA distribution.

Table 4: Cumulative GPA Correlations.

	features	r	p-value
automatic sensing	activity pre-slope	0.418	0.022
	activity post-slope	-0.449	0.015
	activity day pre-slope	0.477	0.008
	activity day post-slope	-0.391	0.036
	activity night pre-slope	0.427	0.019
	activity night post-slope	-0.411	0.027
	conversation duration post-slope	0.443	0.016
	conversation duration night post-slope	0.407	0.028
	conversation duration evening post-slope	0.368	0.050
	conversation freq night breakpoint	0.641	< 0.001
	conversation freq evening breakpoint	0.498	0.005
	indoor mobility term-slope	-0.387	0.035
	indoor mobility pre-slope	0.425	0.019
	indoor mobility post-slope	-0.426	0.021
	indoor mobility night term-slope	-0.396	0.031
	indoor mobility night pre-slope	0.433	0.017
	indoor mobility night post-slope	-0.448	0.015
	indoor mobility day post-slope	-0.386	0.039
	class attendance pre-slope	0.470	0.009
	study duration mean	0.518	0.003
	study focus activity mean	0.430	0.018
	study focus activity pre-slope	-0.372	0.043
	study focus audio mean	0.380	0.038
	study focus audio post-slope	-0.548	0.002
surveys	PHQ-9 depression scale (post)	-0.470	0.027
	conscientiousness	0.551	0.004
	neuroticism	-0.423	0.035

the night epoch and $r = 0.498, p = 0.005$ for the evening epoch).

In terms of the psychological features from pre-post survey data, we find the PHQ-9 [23, 24, 37] score negatively correlates with GPA ($r = -0.470, p = 0.027$), meaning that students who are more depressed tend to have lower grades. Students who are more conscientious from the Big 5 [20] survey tend to have higher GPAs ($r = 0.551, p = 0.004$) and students who tend to be more neurotic are more likely to have lower GPAs ($r = -0.423, p = 0.035$).

Prediction Analysis

In this section, we present a simple model that can predict GPA. We use linear regression with lasso regularization to identify non-redundant predictors among a large number of input features. Theses predictors include a combination of automatic sensing time series behavioral data (i.e., conversational and study features), EMA time series data (e.g., positive affect and stress), mental health data (i.e., depression), and personality data (i.e., conscientiousness).

Predictive Model. Predicting GPA is a regression problem; that is, predicting an outcome variable (i.e., GPA) from a set of input predictors (i.e., features). We evaluate various regression models such as regularized linear regression, regression trees, and support vector regression using cross-validation. We select the *Lasso* (Least Absolute Shrinkage and Selection Operator) [40] regularized linear regression model as our predictive model. Lasso is a method used in linear regression; that is, Lasso minimizes the sum of squared errors, with a bound on the sum of the absolute values of the coefficients. Considering we have a large number of features, collinearity needs to be addressed. There are two categories of methods that address collinearity: feature selection and feature transformation. Lasso regularization is one of the feature selection methods.

Lasso solves the following optimization problem:

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right)$$

where N is the number of observations; y_i is the ground truth of observation i; x_i is the p degree feature vector at observation i; λ is a nonnegative regularization parameter, which controls the number of nonzero components of β (i.e., number of the selected features); β_0 is the intercept; and β is the weight vector. The regularization parameter λ is selected using cross-validation. The optimization problem is essentially to minimize the mean square error $\frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2$ of fitting while keeping the model as simple as possible (i.e., select a minimal number of features to avoid overfitting). Thus, *Lasso* automatically selects more relevant features (i.e., predictors) and discards redundant features to avoid overfitting.

Evaluation Metric. We use the mean absolute errors (MAE), the coefficient of determination (R^2) [7], and Pearson correlation to measure the performance of outcome prediction. MAE measures how close predictions are to the outcomes. The mean absolute error is given by MAE $= \frac{1}{n}\sum_{i=1}^{N}|y_i - \beta_0 - x_i^T\beta|$. Smaller MAE is preferred because it indicates that the predictions are closer to the ground truth. R^2 is a another statistic that measures the goodness of fit of a model and indicates how much of the variance our model explains. R^2 ranges from 0 to 1, where 1 indicates that the model perfectly fits the data. R^2 can be seen to be related to the unexplained variance where $R^2 = 0$ if the feature vector X tells us nothing about the outcome. We use Pearson correlation to measure the linear relations between the ground truth and the predictive outcome.

Table 5: Lasso Selected GPA Predictors and Weights.

	features	weight
sensing	conversation duration night breakpoint	0.3467
	conversation duration evening term-slope	-0.6100
	study duration	0.0728
EMA	positive affect	0.0930
	positive affect post-slope	-0.1215
	stress term-slope	-2.6832
survey	conscientiousness	0.0449

GPA Prediction Results. We apply leave-one-subject-out cross validation [21] to determine the parameters for *Lasso* and the weights for each feature. In order to make the weight regularization work properly, each feature is scaled within the range $[0, 1]$. Selected features have non-zero weights. The MAE of our predicted cumulative GPA is 0.179, indicating that the predictions are within ± 0.179 of the groundtruth. The R^2 is 0.559, which indicates that the features can explain 55.9% of the GPA variance. The predicted GPA strongly correlates with the ground truth with $r = 0.81$ and $p < 0.001$, which further indicates that our predictions can capture outcome differences using the given features.

Table 5 shows the selected features to predict the cumulative GPAs and their weights. Interestingly, *lasso* selects a single long term measure (i.e., conscientious personality trait), two self-report time series features (i.e., affect and stress), and three automatic sensing data behaviors (i.e., conversational and study behavior). The weights indicate the strength of the predictors. Students who have better GPAs are more conscientious, study more, experience positive moods (e.g., joy, interest, alertness) across the term but register a drop in positive affect after the midterm point, experience lower levels of stress as the term progresses, are less social in terms of conversations during the evening period between 6-12 pm, and experience later change (i.e., a behavioral breakpoint) in their conversation duration pattern. In the case of spring term GPA, *lasso* does not select features for prediction. Instead, it chooses to use the intercept alone to predict the spring term GPA outcomes; that is, it chooses a single value 3.40 to predict the spring GPA for all the students with MAE $= 0.53$ and median absolute error of 0.38. We believe this is due to the skewness of the spring term GPAs, as shown in Figure 3(c). Note that average spring term GPA is 3.3306. The intercept *lasso* selected is close to the average GPA but adjusted to the skewness.

DISCUSSION

In this section, we contextualize our findings with regard to the existing literature on the connections between academic performance and students' automatically sensed behaviors, academic-related behaviors, personality, affect, stress, and lifestyle.

We found a number of behavioral change patterns that students experience (i.e., slopes and breakpoints) significantly correlate with academic performance. Previous work [43] studied the level of behaviors or averages across a 10-week term, such as physical activity levels and sociability levels. However, this fails to capture the individual differences among students. It fails to recognize that different people may have different behavioral baselines. For example, extroverts socialize more whereas introverts socialize less because socialization provides more joy to extroverts, thus extroverts and introverts who have the same academic outcome may have different behavior levels. In addition, behavior level fails to capture how behavior changes occurs overtime. By modeling behavioral change using behavioral slopes and breakpoints, we get insights into how different students react to events in their lives albeit social commitments, academic workload, or other hidden triggers that impact students (e.g., stressors or pressures). In this paper, we found that time series analysis allows us to quantify individual behavioral differences over a large timescale of 10 weeks. We also found that slopes and breakpoints correlate with performance and in some cases are strong predictors, as discussed in the Prediction Analysis Section.

In terms of sociability, our results showed that changes in students' conversation durations were significant predictors of performance. Specifically, students who showed change in their conversation durations later in the term for the night epoch had higher GPAs. Students who showed a decrease in their conversation durations during the evening epoch throughout the term also had higher GPAs. These findings are novel and extend previous work that demonstrates a relationship between academic performance and social involvement [31]. Our findings contribute to this body of work by suggesting that changes in students' sociability patterns are important predictors of academic performance. For example, our results suggest that students who change their night time socializing durations later in the term performed better, compared to those who change their night time socializing earlier in the term. Additionally, students who decrease their evening socializing durations during the term perform better, compared to students who increase their evening socializing durations during the term. We suspect that these students may be preparing for their examinations and focusing on other tasks during the evening (e.g., studying), which could contribute to the observed decreases in ambient conversation duration. In addition, our results are consistent with previous research that found greater student outings at night to be associated with lower performance [18].

Turning to academic-related behaviors, our results showed that study duration was a significant predictor of performance. More specifically, students with longer average study durations had higher GPAs at the end of the term, compared to students with shorter study durations. This finding is consistent with research that found academic-related skills (e.g., study skills and habits) to be associated with higher GPAs [31]. Our results extend this work by going beyond self-reported study habits to show that unobtrusively measured studying habits (e.g., via WiFi and GPS) can also predict student performance. In contrast to previous research, we did not find class attendance to be a significant predictor of performance, and we did not observe simple correlations between class attendance and GPAs as other studies have suggested [12]. After inspecting the distribution of the students grades and attendance as shown in Figure 3, we find students who have higher GPAs have either a high or low attendance rate, whereas students who have medium to lower GPAs have a medium attendance rate. The data shows for some high academic performers, attending lectures or not does not affect their grades. We believe students' attendance is determined by the classes they take. Since all of them take at least one programming class, high achievers may not need to attend lectures to perform well.

In terms of personality, our results showed that conscientiousness was a significant predictor of performance, such that students higher in conscientiousness had higher GPAs compared to students lower in conscientiousness. This finding is consistent with psychological research that examines the relationship between self-reported personality traits and academic performance of college students [28]. Although neuroticism was not a significant predictor of students' GPAs, an examination of the simple correlations between the traits and GPAs shows that neuroticism was negatively associated with cumulative GPA performance ($r = -0.42$), suggesting that students who are higher in neuroticism have lower GPAs.

In terms of affect, our results showed that positive affect levels and change were significant predictors of performance. Specifically, students with higher average levels of positive affect had higher GPAs at the end of the term, compared to students with lower average levels. Students with decreasing positive affect after the midterm point also had higher GPAs, compared to students with increasing positive affect. These findings are consistent with previous studies that demonstrate a relationship between greater positive affect and higher grades and GPAs [32]. Interestingly, previous studies have found that increase in negative affect during the second half of a semester is associated with lower grades and GPAs [32]. However, we found that decreases in positive affect during the second half of the term were associated with higher GPAs. We suspect that students who focused on their academic-related tasks and performance during the second half of the term are less likely to report feeling great positive affect (e.g., excited, enthusiastic), compared to students who did not .

In terms of stress, our results showed that the change in stress during the term was a significant predictor of performance, such that students with decreases in stress levels throughout the term had higher GPAs at the end of the term. This finding is consistent with previous research that finds a negative relationship between student stress and GPAs (e.g., [30]). We suspect that students who performed well in their classes became less stressed as the semester progressed, compared to students who performed poorly in their classes. This is consistent with previous research that found greater stress during the end of the semester to be associated with lower GPAs [29].

We are the first to the best of our knowledge that has used automatic sensing data from smartphones and time series EMAs to predict GPA. The predicted GPA strongly correlates with the groundtruth with $r = 0.81$ and $p < 0.001$, MAE is 0.179, R^2 is 0.559. We predict GPA without using any priors such as SAT, IQ test results, or knowing students' grades during the term. While a student in day-to-day life would likely measure their success via assignment grades and midterm performance, our results show that there are a number of other predictors of academic success to consider. Our prediction model indicates that students getting better grades are more conscientious, study more, experiences positive moods across the term but register a drop in positive affect after the midterm point, experience lower levels of stress as the term progresses, are less social in terms of conversations during the evening, and experience change in their conversation duration pattern later in the term. The correlations and prediction model discussed in this paper naturally lead to a consideration of potential interventions to improve academic performance. We hypothesize that our work could serve as a catalyst for new forms of real-time interventions to help under-performing students improve their academic performance.

We also recognize the limitations of our work. Although the dataset [43] is large, rich, and deep, the number of students in the study is small (N=30). Such a small dataset is limiting because we cannot use more sophisticated predictive models or features because it may lead to overfitting. We see from the predictive results that *Lasso* selects only 7 features from 193 features despite that we found many more correlations. Next, Dartmouth is an Ivy league liberal arts college. Its undergraduates are among the top high school performers. Therefore, our sample is skewed to high performers with good GPAs. Finally, while the students in the sample were not all computer science majors they all took one class in common [43]: Android programming. The samples therefore could be biased to science students and do not represent the larger cross section of students found in liberal arts, for example. We believe that a larger scale study with more diverse college students across different universities would present better samples for our study. Such a large scale, cross institutional study would offer more diverse and representative samples allowing us to refine and revalidate our predictive model accordingly.

CONCLUSION

The SmartGPA study has shown that there a number of significant correlations between GPA and a number of behaviors automatically inferred from smartphone sensing data. We also presented a number of novel automatic sensing methods for assessing the study and social behavior of students, including, partying instances and duration, and study duration and focus. In our previous StudentLife study [43], we used simple averages of all student behaviors over the term and presented a number of correlations with performance. The SmartGPA study goes much deeper in our analysis of academic performance and proposes time series analysis of each student's data streams to best understand individuals' differences between high and low performers. As part of the analysis we proposed novel methods to assess behavioral changes experienced by students over the 10-week term – that is, we proposed behavioral slopes and breakpoints to capture changing behaviors. Furthermore, we proposed a simple predictive model that use linear regression with *lasso* regularization on a number of performance predictors. The predicted GPAs are within ± 0.179 of the groundtruth. Our results open the way for novel interventions to improve academic performance.

REFERENCES

1. Dartmouth bans hard liquor in effort to end high-risk and harmful behavior. `http://college.usatoday.com/2015/01/31/dartmouth-bans-hard-liquor-in-effort-to-end-high-risk-and-harmful-behavior/`.
2. Dartmouth College Weekly Schedule Diagram. `http://oracle-www.dartmouth.edu/dart/groucho/timetabl.diagram`.
3. Moving dartmouth forward. `http://www.dartmouth.edu/~president/forward/`.
4. Piazza. `https://piazza.com/`.
5. Regulations for Graduate Study. `http://www.dartmouth.edu/~regarchive/catalog/regulations04/grad.html/`.
6. BUYSSE, D. J., REYNOLDS III, C. F., MONK, T. H., BERMAN, S. R., AND KUPFER, D. J. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research 28*, 2 (1989), 193–213.
7. CAMERON, A. C., AND WINDMEIJER, F. A. R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics 14*, 2 (1996), 209–220.
8. CHAMORRO-PREMUZIC, T., AND FURNHAM, A. Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality 37*, 4 (2003), 319 – 338.
9. CHEN, F., WANG, R., ZHOU, X., AND CAMPBELL, A. T. My smartphone knows i am hungry. In *Proceedings of the 2014 Workshop on Physical Analytics* (New York, NY, USA, 2014), WPA '14, ACM, pp. 9–14.
10. COHEN, S., KAMARCK, T., AND MERMELSTEIN, R. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396.
11. CONARD, M. A. Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality 40*, 3 (2006), 339 – 346.
12. CREDÉ, M., ROCH, S. G., AND KIESZCZYNKA, U. M. Class attendance in college a meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research 80*, 2 (2010), 272–295.
13. DEBERARD, M. S., SPIELMANS, G., AND JULKA, D. Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College student journal 38*, 1 (2004), 66–80.
14. DIENER, E., WIRTZ, D., TOV, W., KIM-PRIETO, C., CHOI, D.-W., OISHI, S., AND BISWAS-DIENER, R. New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research 97*, 2 (2010), 143–156.
15. FAUSETT, L., AND ELWASIF, W. Predicting performance from test scores using backpropagation and counterpropagation. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on* (Jun 1994), vol. 5, pp. 3398–3402 vol.5.
16. FURNHAM, A., CHAMORRO-PREMUZIC, T., AND MCDOUGALL, F. Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences 14*, 1 (2002), 47 – 64.
17. GILBERT, S. P., AND WEAVER, C. C. Sleep quality and academic performance in university students: A wake-up call for college psychologists. *Journal of College Student Psychotherapy 24*, 4 (2010), 295–306.
18. GOMES, A. A., TAVARES, J., AND DE AZEVEDO, M. H. P. Sleep and academic performance in undergraduates: a multi-measure, multi-predictor approach. *Chronobiology International 28*, 9 (2011), 786–801.
19. GONZALEZ, E. C., HERNANDEZ, E. C., COLTRANE, A. K., AND MANCERA, J. M. The correlation between physical activity and grade point average for health science graduate students. *OTJR: occupation, participation and health 34*, 3 (2004), 160–167.
20. JOHN, O. P., AND SRIVASTAVA, S. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research 2* (1999), 102–138.
21. KOHAVI, R., ET AL. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (1995), vol. 14, pp. 1137–1145.
22. KOTSIANTIS, S., AND PINTELAS, P. Predicting students marks in hellenic open university. In *Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on* (July 2005), pp. 664–668.
23. KROENKE, K., AND SPITZER, R. L. The phq-9: a new depression diagnostic and severity measure. *Psychiatric Annals 32*, 9 (2002), 509–515.
24. KROENKE, K., SPITZER, R. L., AND WILLIAMS, J. B. The phq-9. *Journal of general internal medicine 16*, 9 (2001), 606–613.
25. MARTINEZ, D. Predicting student outcomes using discriminant function analysis.
26. PAULHUS, D. L., AND VAZIRE, S. The self-report method. *Handbook of research methods in personality psychology* (2007), 224–239.
27. POLLAK, J. P., ADAMS, P., AND GAY, G. PAM: a photographic affect meter for frequent, in situ measurement of affect. In *Proc. of SIGCHI* (2011).
28. POROPAT, A. E. A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin 135*, 2 (2009), 322.

29. RAFIDAH, K., AZIZAH, A., NORZAIDI, M. D., CHONG, S. C., SALWANI, M. I., AND NORAINI, I. Stress and academic performance: Empirical evidence from university students. *Academy of Educational Leadership Journal 13*, 1 (2009), 37.

30. RETTINGER, V. W. The relationship between physical activity, stress, and academic performance. *Doctoral dissertation*.

31. ROBBINS, S. B., LAUVER, K., LE, H., DAVIS, D., LANGLEY, R., AND CARLSTROM, A. Do psychosocial and study skill factors predict college outcomes? a meta-analysis. *Psychological bulletin 130*, 2 (2004), 261.

32. ROGATEN, J., MONETA, G., AND SPADA, M. Academic performance as a function of approaches to studying and affect in studying. *Journal of Happiness Studies 14*, 6 (2013), 1751–1763.

33. ROMERO, C., ESPEJO, P. G., ZAFRA, A., ROMERO, J. R., AND VENTURA, S. Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education 21*, 1 (2013), 135–146.

34. ROSS, S. E., NIEBLING, B. C., AND HECKERT, T. M. Sources of stress among college students. *Social psychology 61*, 5 (1999), 841–846.

35. RUSSELL, D. W. UCLA loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of personality assessment 66*, 1 (1996), 20–40.

36. SCHWARZ, G., ET AL. Estimating the dimension of a model. *The annals of statistics 6*, 2 (1978), 461–464.

37. SPITZER, R. L., KROENKE, K., WILLIAMS, J. B., GROUP, P. H. Q. P. C. S., ET AL. Validation and utility of a self-report version of prime-md: the phq primary care study. *Jama 282*, 18 (1999), 1737–1744.

38. TAMHANE, A., IKBAL, S., SENGUPTA, B., DUGGIRALA, M., AND APPLETON, J. Predicting student risks through longitudinal analysis. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2014), KDD '14, ACM, pp. 1544–1552.

39. TAYLOR, D. J., VATTHAUER, K. E., BRAMOWETH, A. D., RUGGERO, C., AND ROANE, B. The role of sleep in predicting college academic performance: Is it a unique predictor? *Behavioral sleep medicine 11*, 3 (2013), 159–172.

40. TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*, 1 (1996), pp. 267–288.

41. TROCKEL, M. T., BARNES, M. D., AND EGGET, D. L. Health-related variables and academic performance among first-year college students: implications for sleep and other behaviors. *Journal of American college health 49*, 3 (2000), 125–131.

42. WALD, A., MUENNIG, P. A., O'CONNELL, K. A., AND GARBER, C. E. Associations between healthy lifestyle behaviors and academic performance in us undergraduates: a secondary analysis of the american college health association's national college health assessment ii. *American Journal of Health Promotion 28*, 5 (2014), 298–305.

43. WANG, R., CHEN, F., CHEN, Z., LI, T., HARARI, G., TIGNOR, S., ZHOU, X., BEN-ZEEV, D., AND CAMPBELL, A. T. Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2014), UbiComp '14, ACM, pp. 3–14.

44. WATANABE, J.-I., MATSUDA, S., AND YANO, K. Using wearable sensor badges to improve scholastic performance. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (2013), ACM, pp. 139–142.

45. WATANABE, J.-I., YANO, K., AND MATSUDA, S. Relationship between physical behaviors of students and their scholastic performance. In *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)* (2013), IEEE, pp. 170–177.

46. WATSON, D., CLARK, L. A., AND TELLEGEN, A. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology 54*, 6 (1988), 1063.

47. XU, R., WUNSCH, D., ET AL. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on 16*, 3 (2005), 645–678.

48. ZAFRA, A., ROMERO, C., AND VENTURA, S. Multiple instance learning for classifying students in learning management systems. *Expert Systems with Applications 38*, 12 (2011), 15020 – 15031.