Exploring Compliance: Observations from a Large Scale Fitbit Study

Louis Faust, Rachael Purta, David Hachen, Aaron Striegel, Christian Poellabauer, Omar Lizardo, Nitesh V. Chawla
University of Notre Dame
Notre Dame, Indiana, USA
{lfaust,rpurta,dhachen,striegel,cpoellab,olizardo,nchawla}@nd.edu

ABSTRACT

Universities often draw from their student body when conducting human subject studies. Unfortunately, as with any longitudinal human studies project, data quality problems arise from student's waning compliance to the study. While incentive mechanisms may be employed to boost student compliance, such systems may not encourage all participants in the same manner. This paper coupled student's compliance rates with other personal data collected via Fitbits, smartphones, and surveys. Machine learning algorithms were then employed to explore factors that influence compliance. With such insight, universities may target groups in their studies who are more likely to become non-compliant and implement preventative strategies such as tailoring their incentive mechanisms to accommodate a diverse population. In doing so, data quality problems stemming from failing compliance can be minimized.

CCS CONCEPTS

•Human-centered computing →Empirical studies in ubiquitous and mobile computing;

KEYWORDS

mobile sensing; machine learning; user studies; social aspects;

ACM Reference format:
Louis Faust, Rachael Purta, David Hachen, Aaron Striegel, Christian Poellabauer, Omar Lizardo, Nitesh V. Chawla. 2016. Exploring Compliance: Observations from a Large Scale Fitbit Study. In *Proceedings of The 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, Pittsburgh, PA, USA, April 21 2017 (SocialSens'17),* 6 pages.
DOI: http://dx.doi.org/10.1145/3055601.3055608

1 INTRODUCTION

Walking onto any campus, one is bound to come across a bulletin board with flyers seeking participants for various studies, as universities often target their student body for human subject studies. This can be particularly useful for longitudinal studies as the students are in close proximity to researchers, allowing for easier interaction and communication if necessary. However, as with any long-term study, compliance on behalf of the participants may dwindle as the novelty of the study wears off [9, 10].

Recent studies involving mobile sensors such as Fitbits have found ways to calculate compliance to a certain degree [7]. The NetHealth study provided students with Fitbit Charge HRs to monitor their health activity. Students were asked to keep their Fitbits on as much as possible and sync their data on a regular basis. From this, compliance scores were calculated for each student based on the Fitbit heart rate monitor. This paper discusses pairing these scores with other personal data from the study and then employing machine learning algorithms to explore factors that influence a student's level of compliance. With the ability to identify subgroups which are less likely to remain compliant, researchers will be able to implement preventative strategies such as tailoring their incentive mechanisms to accommodate these students. Doing so will potentially prevent data loss due to non-compliance.

2 RELATED WORK

Maintaining strong compliance on behalf of participants is a common problem for longitudinal studies [5, 9, 10]. Recent studies [5, 9] point to issues ranging from study design to incidental constraints stemming from technical quirks in tracking devices.

The Fitbit Zip study's [5] purpose was to identify problems in longitudinal studies using physical tracking devices. A short pilot was performed involving four participants, but technical problems arose when the study was extended to 16 weeks and expanded to 50 users. Thirty participants experienced technical issues with the Fitbits or lost the device altogether. The Fitbits were able to hold 30 days worth of data before needing to be synced. Therefore, the researchers scheduled meetings with the participants every 28 days to sync their Fitbits and conduct interviews. However, participants generally ignored the meeting requests, causing their Fitbits to overwrite data due to the postponed syncing.

The Zip study demonstrates the importance of convenience on behalf of the participants for maintaining higher compliance; but convenience does not ensure compliance. The NetSense study [9] relied on monitoring a variety of communication media (Facebook, SMS, e-mail, phone calls, etc.) unobtrusively to monitor the capacity of individuals to make and keep friends in a novel context (high school students transitioning to college). This was done by giving incoming freshmen smartphones which could be monitored. Students were asked to complete entrance surveys covering personality aspects and current social network status and from there, on-going surveys were distributed online and via an app on the phone. This allowed all participation required by the study to be managed directly from students phones. In spite of this apparently

SocialSens'17, Pittsburgh, PA, USA
© 2017 ACM. 978-1-4503-4977-2/17/04…$15.00
DOI: http://dx.doi.org/10.1145/3055601.3055608

convenient setup, compliance dropped over the course of the study as students decreased their use of the the assigned phone and became less responsive to completing surveys. While the authors attribute this to students becoming busier (moving to sophomore year) and becoming less interested in remaining compliant, they also note the absence of an ongoing incentive mechanism as a factor contributing to the decline.

Incentive mechanisms are often employed as means of maintaining compliance throughout longitudinal studies. Researchers [6] surveyed a variety of incentive mechanisms geared toward crowd-sensing. Incentives ranged from monetary to volunteer-based to self-beneficial. The authors found several components were required for an incentives mechanism to be considered successful. Among these components were economic feasibility, how the mechanism encourages high-quality data collection, ensuring equal opportunity to all participants, scalability, area coverage, and management. No single incentive mechanism was found to fit all scenarios on the side of the researchers or participants. The authors believed an ideal system would consider each user's profile to determine the most appropriate incentive mechanism per user to maximize compliance. NetHealth employs a static monetary model of $35 a month to those who remain compliant and at one point used a dynamic monetary model, offering compensation based on student's level of compliance. As a result, participants compliance scores reflect natural variations in behavior along with a responsiveness to these particular incentive models.

3 DATA

The NetHealth study involves an ongoing collection of survey, phone, and Fitbit data on approximately 700 first-year students who entered the university in the Fall of 2015. Recruitment was done via an interest survey in June 2015 and solicitations made through e-mail and a Facebook page. All students received a Fitbit Charge HR. Some received their Fitbit before arriving on campus, others after arrival, and some early in the Spring 2016 semester. Surveys administered to NetHealth participants include self-report questionnaires covering different aspects of individuals demographics, mental and physical health, and relations with other students.

- Educational Background: Type of high school attended; average grades; advanced placement courses; extra-circulars; and fields of interest.
- Personality: Big-Five personality test; self-efficacy assessments in areas of relationships, health, exercise, and diet.
- Family: Parents' status; siblings; family's educational and economic status; religion.
- Activities: Hobbies; favorite exercise activities; use of tobacco, alcohol, and drugs.
- Physical Assessment: Level of satisfaction in terms of happiness, general health, body image, potential disabilities, and current medications.
- Mental Health: Screening tests for depression and anxiety; recent major life changes.
- Sleep: General sleep habits such as average time going to bed, waking up, and hours of sleep; troubles falling asleep; how they felt after waking up.
- Attitudes: Students political affiliation; stance on popular political topics such as the legalization of marijuana.
- Technology: Devices students use to communicate with friends and family; devices owned; usage of social media applications.
- Demographics: Gender; race; whether they are a US citizen.

Table 1: Class distribution of students at an 80% compliance threshold

Class	Frequency
Compliant	213
Non-compliant	287

The phone data maps student's social networks to determine each student's social ties and number of people they are regularly in contact with based on SMS messages and voice calls. This information is used to approximate the degree, or number of friends each student had, and the average degree of each student's friends.

Fitbit metrics include, among other things, heart rate, active minutes, steps, and minutes asleep. Heart rate ranges are based on a percentage of a maximum heart rate that Fitbit estimates for each user. For instance, *Out of Zone*, or what we refer to as *Low Range* is the total number of minutes the users heart rate was below 50% of their estimated maximum heart rate, while *High Range* is an aggregate of the Fitbit's *Fat Burn*, *Cardio*, and *Peak* ranges, beginning with a rate above 50% of a user's estimated maximum [3]. Active minutes are calculated using metabolic equivalents (METs) as these measure energy expenditure and are weight agnostic. Fitbit measures active minutes by periods of 10 minutes or more for which the user maintains a level at or above 3 METs [2].

Compliance rates are derived from the Fitbit data by counting the number of minutes throughout the day that the Fitbit detects a heart rate, as Fitbit will only record a heart rate if it's on the wrist. This sum is divided by the total minutes in a day (1440) to produce a daily compliance percentage. These daily percentages were then averaged into an overall compliance score for each student based on the 2015-2016 academic year and following Fall semester of 2016 (42 weeks total). Fall, Winter, Spring and Summer breaks were removed from consideration due to findings that compliance issues and data loss were most severe during these breaks, most likely caused by Fitbit syncing issues, broken devices, or students forgetting to sync [7].

Each student was classified as *compliant* or *non-compliant* based on whether or not the student's average compliance rate was above or below an 80% threshold. Table 1 summarizes the class distribution, with *non-compliant* students accounting for about 15% more of the population. For consistency, we only examined the initial 500 students who entered in the Summer and Fall of 2015. Based on Figure 1, we note a fair amount of variance in scores falling below the 80% mark, suggesting these students are still participating to some degree instead of dropping out entirely. With each student classified, machine learning algorithms were applied to parse the data and determine factors influencing a student's compliance.

4 METHODS

The NetHealth repository contains data covering a diverse set of traits and behaviors contributing to a user's personal profile. Given

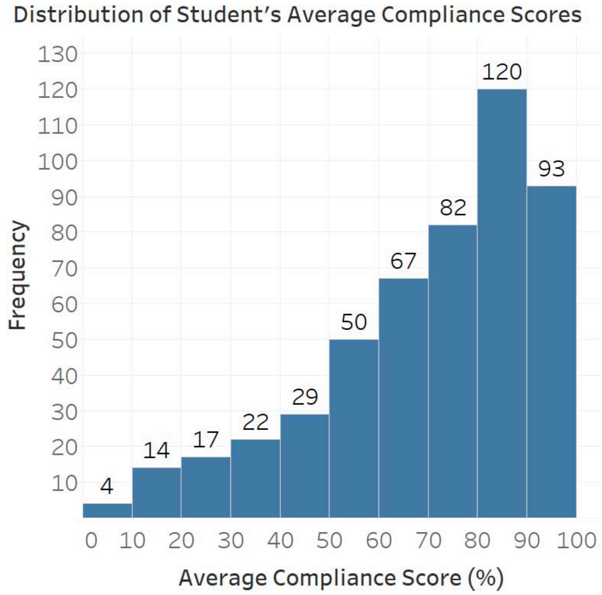

Figure 1: Histogram of all student's average compliance scores across the entire study

this wide data set, feature selection methods were used to narrow the feature space. The first step was to use random forests [1], which ranked variables based on their contribution to classifying a student. A random forest was fitted using 100 trees with a Gini impurity criterion, which ranked the features across twenty-five runs, fitting a new random forest on each iteration. Variables that did not contribute to the model were given scores of zero and removed from consideration. A new series of twenty-five runs was then performed on the remaining variables. This process continued until all variables remaining in the model had a score of greater than zero.

Having removed a large portion of noise from the data, a recursive feature elimination (RFE) model [4] was run to determine the optimal set of remaining features to use. Using the same random forest classifier with 10-fold cross validation, two separate instances were run, with one scoring the model by classification accuracy and the other by area under an ROC curve (AUROC). The RFE models produced optimal models with over 100 features, however, as Figures 2 and 3 show, the majority of this accuracy is accounted for with less than twenty features. This subset of features showed three categories: Personality traits (Big-Five), Fitbit metrics, and self-efficacy scales. The self-efficacy scales were removed from further consideration given the category was considered a less stable aspect of a person's profile. Specifics for these categories are provided in Table 2.

Decision trees were then drafted for each category to see how the features affected compliance rates, as their hierarchical nature visualizes which features are the strongest predictors and as branches are followed down to the leaf nodes, subgroups become more detailed. Visualizations showing compliance trends over the study were also created to see how factors increased or decreased compliance over time.

Table 2: Top Features Derived from RFE by Category

Personality Traits	Fitbit Metrics (Averages)
Extraversion	Steps
Neuroticism	Minutes Asleep
Agreeableness	Sedentary Minutes
Openness	Active Minutes
Conscientiousness	Low Heart Rate Minutes
	High Heart Rate Minutes

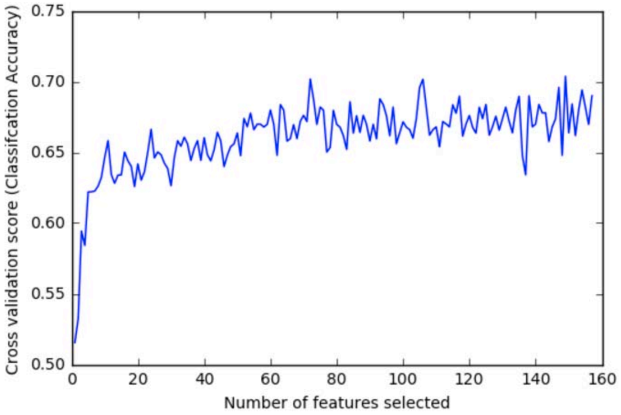

Figure 2: RFE plot showing the optimal number of features for the model based on classification accuracy.

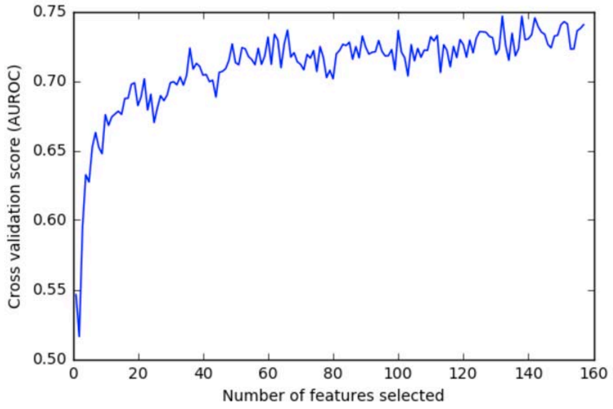

Figure 3: RFE plot showing the optimal number of features for the model based on AUROC.

5 ANALYSIS

5.1 Decision Trees

To investigate the major factors (personality and Fitbit metrics), decision trees [8] were drafted to explore which aspects of personality traits and Fitbit metrics affected whether a student was compliant or not. These trees were constructed using only personality traits and Fitbit metrics respectively with a Gini impurity splitter. Decision trees were also split via information gain, however, this did not change the structure of the trees.

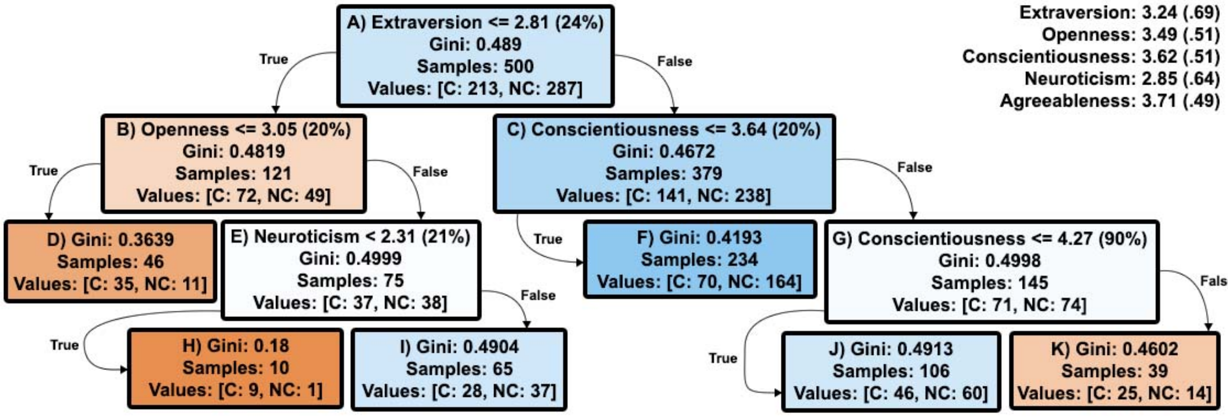

Figure 4: Decision tree showing personality traits as splitting factors for classifying students as *compliant* or *non-compliant*. Averages and standard deviations for each metric are shown in the top right, while each node shows the splitting threshold as a score and percentile, gini impurity, number of samples, and students of that sample belonging to *compliant* and *non-compliant* classes respectively ([*compliant, non-compliant*]).

The personality traits decision tree in Fig. 4 breaks compliance factors into several paths, leading to the majority of students being classified as *compliant* or *non-compliant*. Each level of a personality trait is based on a score of 1-5 calculated from the Big Five Personality Assessment.

The tree first separates students by level of extraversion, splitting at the 24th percentile. The distributions in nodes B and C show students who are more introverted ([A, B]) with stronger signs of compliance than extroverts ([A, C]). The tree then splits these more introverted students by their Openness score, with students who are less open ([A, B, D]), showing stronger compliance. Students who are more open ([A, B, E]), however, show a near even distribution between the two classes. This group is further split on their levels of neuroticism, with high compliance favoring those with low neuroticism scores ([A, B, E, H]) and those with average to high scores ([A, B, E, I]) showing only minor signs of non-compliance. Focusing on the other half of the tree, extroverts are split by their levels of conscientiousness with those in the 20th percentile ([A, C, F]) being predominantly non-compliant. Those with higher levels of conscientiousness ([A, C, G]) are fairly balanced and are separated again by the same factor, this time splitting at the 90th percentile. Those above the threshold ([A, C, G, K]) show moderate signs of compliance, while those below ([A, C, G, J]) show moderate signs of non-compliance.

The decision tree for the Fitbit metrics in Fig. 5 proved difficult to parse due to the overlapping nature of the metrics, i.e. low/high heart range minutes and sedentary/active minutes. This left only a few factors to split on: minutes asleep, steps, heart range minutes, and active minutes, with minutes asleep and active minutes providing the only significant splits. The tree showed that students who were more active (fewer sedentary minutes) ([A, C, D]) had higher compliance rates than those less active ([A, C, E]) and students who slept longer had higher compliance rates than those who slept less ([A, B]).

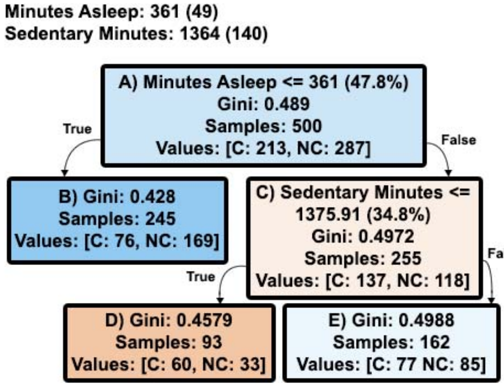

Figure 5: Decision tree showing Fitbit metrics as splitting factors for classifying students as *compliant* or *non-compliant*. Node details are presented in the same structure as Figure 4.

5.2 Compliance Trends Over Study

To see how these factors influenced compliance throughout the study, daily compliance scores were computed for each student and then averaged across all students for each day. These averages were then broken down into subgroups based on median-splits for each personality trait score ranging from 1-5. The Fall 2015 semester uses personality scores taken from the first survey students completed in early Summer or Fall of 2015, while the Spring 2016 and Fall 2016 semesters use scores taken from a later survey completed in early 2016.

Fig 6 shows compliance trends for each semester based on level of extraversion. Across the three semesters, those more introverted showed stronger signs of compliance overall, however, compliance decreased at a rate similar to those who are more extroverted. Only in the third semester is there a significant change with both groups starting at roughly the same score with introverts increasing their compliance rates while extroverts declined.

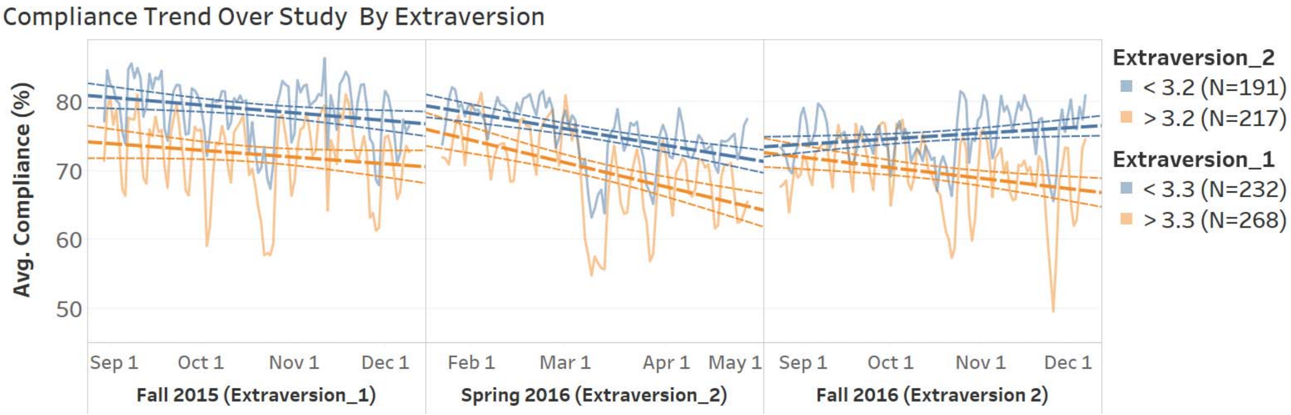

Figure 6: Average daily compliance scores by level of extraversion across the Fall 2015, Spring 2016 and Fall 2016 semesters.

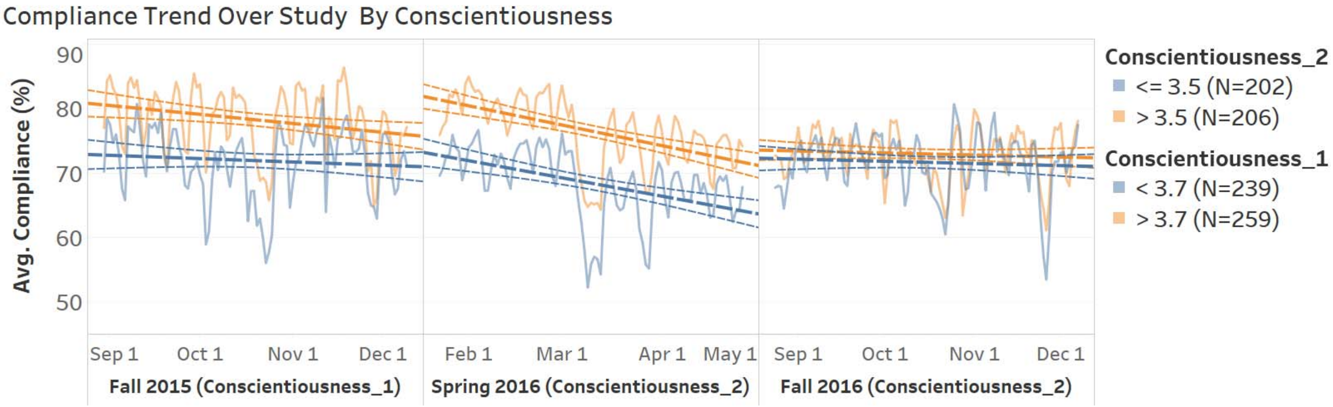

Figure 7: Average daily compliance scores by level of conscientiousness across the Fall 2015, Spring 2016 and Fall 2016 semesters.

Fig 7 shows compliance trends for each semester based on level of conscientiousness. The first two semesters show similar slopes, with Spring 2016 having a stronger decline than Fall 2015, but those more conscientious have higher levels of compliance. The Fall 2016 semester, however, shows almost no difference between the two groups.

Fig 8 shows compliance trends for the semester based on level of openness. The first semester shows little difference between the groups. The Spring 2016 semester has similar slopes between the two groups, however, those who are less open maintained a higher average compliance than those more open. A similar difference appears in the Fall 2016 semester, however, those who were less open had a positive slope while those who were more open had a negative slope.

Compliance trends were not included for neuroticism and agreeableness due to their relatively low significance. Only those with extreme levels of neuroticism differed in compliance. Those below the 7.4 percentile (N=44) had above average compliance while those above the 98.5 percentile (N=24) had below average rates. Agreeableness had a similar trend where only those below the 6.2 percentile (N=31) had below average compliance.

6 DISCUSSION

Based on the feature analysis, twenty features emerged as the strongest factors predicting compliance, those further considered were six Fitbit metrics and five personality traits. The inclusion of almost all the Fitbit metrics raised some initial concerns regarding whether there was a bias present as the classes *compliant* and *non-compliant* were derived from a Fitbit metric. However, because several of the remaining features were non-Fitbit metrics, but still related to health status, this could validate the Fitbit metrics being ranked highly, as they too represent overall health.

As the personality traits decision tree shows, students with higher extraversion and higher openness respectively have lower compliance rates. Because forgetting to wear/sync/or charge the Fitbit is the primary reason for non-compliance, personality traits may cause students to forget about the Fitbit. Extroverted students are more social, a trait which may leave them easier to distract, resulting in forgetting to wear the Fitbit more than an introvert. Students with high openness may be less likely to fall into routines, preventing Fitbit compliance from being a daily concern, or they may quickly grow bored with the study in comparison to those who are less open and can more easily include Fitbit compliance into their daily routine. Since higher openness leads to lower compliance, but lower extraversion leads to higher compliance, it's

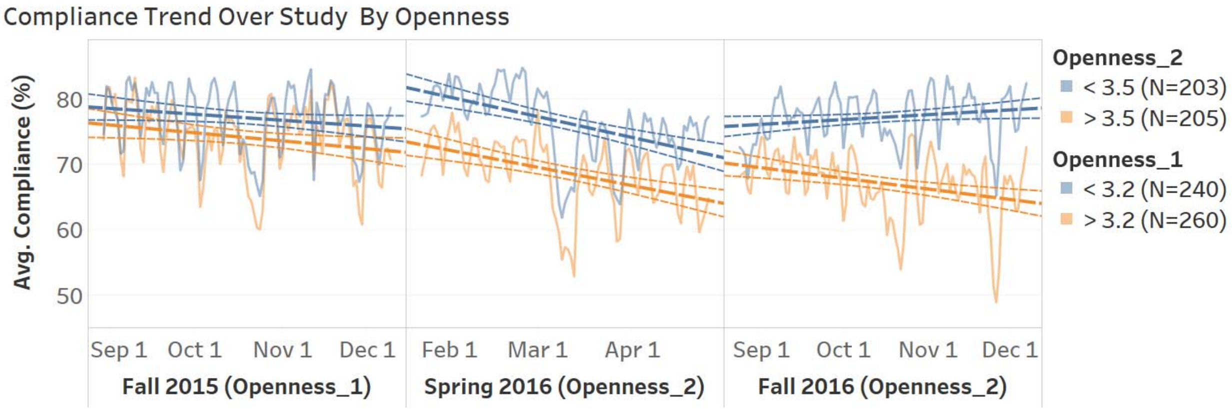

Figure 8: Average daily compliance scores by level of openness across the Fall 2015, Spring 2016 and Fall 2016 semesters.

reasonable to assume a combination of these traits would sway students to compliance or non-compliance. Extroverts with high conscientiousness tend to be more compliant, meaning they remain mindful of staying compliant amidst any distractions. As for neuroticism, only those with much lower levels tended to be more compliant which may result from being less distracted by feelings of anxiety, fear, loneliness, etc.

For the Fitbit decision tree, minutes asleep and sedentary minutes were the only metrics providing significant distinction between *compliant* and *non-compliant* students. Better compliance from sleeping more and being more active could be the result of students who are generally healthier being more interested in wearing their Fitbits. Given the Fitbit is a health tool, it would logically follow that those interested in their health would use their Fitbit more, leading to better compliance scores. This raises another concern in that students with high compliance may not be as interested in wearing the Fitbit for the sake of remaining compliant to the study than they are wearing it for its intended purpose: tracking their personal health. Further investigation is considered in Section 8.

These differing levels of compliance represent natural variations in behavior as well as how students have responded to the incentive models employed by NetHealth. To better understand how groups more likely to become non-compliant can be motivated, alternative incentive mechanisms should be tested to gauge how students in these groups would respond.

7 CONCLUSIONS

Determining factors that influence compliance can aid researchers in targeting those more susceptible to non-compliance and tailoring their incentive mechanisms accordingly. Doing so can minimize data quality issues when a portion of the participants' compliance may start to wane. This paper explored factors influencing compliance of students in a longitudinal Fitbit study by pairing compliance data with other personal factors. Students were classified as *compliant* or *non-compliant* based on average daily compliance scores derived from the amount of time their Fitbit was worn. A feature analysis showed personality traits (in particular, extraversion, openness, and conscientiousness) and Fitbit metrics such as minutes asleep and sedentary minutes to be the dominant factors influencing compliance.

8 FUTURE RESEARCH

The next step is to address the issue of Fitbits acting as a compliance indicator and health information device. Additional factors such as how often students sync their Fitbit and phone data may provide a stronger overall picture of compliance, as students need to sync this data frequently to be considered compliant within the study.

ACKNOWLEDGEMENTS

The research reported in this paper was supported by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health under award number R01HL117757. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32. DOI: http://dx.doi.org/10.1023/A:1010933404324
[2] FitBit. 2016. What are active minutes? (2016). https://help.fitbit.com/articles/en_US/Help_article/1379
[3] FitBit. 2016. What should I know about my heart rate data? (2016). https://help.fitbit.com/articles/en_US/Help_article/1565#zones
[4] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine learning* 46, 1-3 (2002), 389–422.
[5] Daniel Harrison, Paul Marshall, Nadia Berthouze, and Jon Bird. 2014. Tracking physical activity: problems related to running longitudinal studies with commercial devices. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication.* ACM, 699–702.
[6] Luis G. Jaimes, Idalides J. Vergara-Laurens, and Andrew Raij. 2015. A Survey of Incentive Techniques for Mobile Crowd Sensing. *IEEE Internet of Things Journal* 2, 5 (Oct 2015), 370–380. DOI: http://dx.doi.org/10.1109/JIOT.2015.2409151
[7] Rachael Purta, Stephen Mattingly, Lixing Song, Omar Lizardo, David Hachen, Christian Poellabauer, and Aaron Striegel. 2016. Experiences Measuring Sleep and Physical Activity Patterns Across a Large College Cohort with Fitbits. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16).* ACM, New York, NY, USA, 28–35. DOI: http://dx.doi.org/10.1145/2971763.2971767
[8] J. Ross Quinlan. 1993. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
[9] Aaron Striegel, Shu Liu, Lei Meng, Christian Poellabauer, David Hachen, and Omar Lizardo. 2013. Lessons Learned from the Netsense Smartphone Study. *SIGCOMM Comput. Commun. Rev.* 43, 4 (Aug. 2013), 51–56. DOI: http://dx.doi.org/10.1145/2534169.2491171
[10] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. 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 (UbiComp '14).* ACM, New York, NY, USA, 3–14. DOI: http://dx.doi.org/10.1145/2632048.2632054