Analyzing the Mental Health of Engineering Students using Classification and Regression

Melissa Deziel, Dayo Olawo, Lisa Truchon, Lukasz Golab
University of Waterloo, Canada
{madeziel, oaolawo, laatruch, lgolab}@uwaterloo.ca

ABSTRACT

In this paper, we describe a data mining study of the mental health of undergraduate Engineering students in a large Canadian university. We created a survey based on guidelines from the Canadian Mental Health Association, and applied classification and regression algorithms to the collected data. Our results reveal interesting relationships between various aspects of mental health and year of study (first and final year students have lower mental health scores than second-year students), academic program (students in competitive programs have lower overall mental health but higher self-actualization, whereas students in a program with a flexible curriculum had higher overall scores), and gender (female Engineering students tend to have lower scores).

Keywords

Mental health data mining, linear regression, rule mining

1. INTRODUCTION

Mental health affects all facets of daily life and therefore awareness is critical. In particular, the well-being of undergraduate students is imperative as it greatly affects their academic success. In this paper, we advocate the use of data mining to understand the factors affecting mental health. We describe a case study in which we applied regression and classification algorithms to mental health survey data collected in a large Canadian university. We focus on Engineering, a competitive and demanding discipline with a heavy gender bias towards male students.

We conducted an anonymous survey - both online and in person - in which we asked students to rate five aspects of their mental health, as defined by the Canadian Mental Health Association [4]. These five aspects are: Ability to Enjoy Life, Resilience, Balance, Emotional Flexibility and Self-Actualization. Our survey also included questions about potential academic influences on mental health such as year of study, academic program, gender, academic workload, and relationship status. We received over 300 responses in total.

We then applied linear regression and classification algorithms to identify which of the above external influences have the greatest effect on each aspect of mental health. Examining the regression coefficients and classification rules revealed interesting insights into the mental health of Engineering students. We found that the number of hours of homework was the best predictor of overall mental health, followed by year of study. In particular, first-year and final-year students tend to have lower mental health scores while second-year students have the highest scores. We also found that female students in all academic programs and years have lower overall mental health but higher Emotional Flexibility. Another interesting result was that students in highly competitive and challenging programs have lower overall mental health scores but higher Self-Actualization, whereas students in inter-disciplinary programs with a flexible curriculum tend to have higher scores.

The aim of this paper is to illustrate how data mining algorithms can be used to analyze mental health data and to encourage further work on applying machine learning to better understand mental well-being. We discuss related work in Section 2, our methodology in Section 3, and our results in Section 4. We conclude in Section 5 with recommendations arising from this study and suggestions for future work.

2. RELATED WORK

The importance of mental health and well-being in students is exemplified by the large number of studies on this topic. Past research has focused on using surveys to identify factors that affect mental health, but applying machine learning tools to such data has not received much attention. One recent example is Li et al. [3], which examined variables such as ethnicity, gender and age to classify the mental health of Chinese college students into three groups using regression. They found that the strongest predictor of adjustment and severe mental health problems was the level of satisfaction with one's major. In this paper, we consider a wider variety of education-related features, including gender, year of study, academic workload and academic program, and we examine five different aspects of mental health.

Another interesting example of mining mental health data is described in Diederich et al. [1], which used machine learning techniques to identfy mental health issues such as schizophrenia, and mania. Their thesis was that through analyzing data or language and conversations between psychiatrists and patients, they could develop more accurate diagnostic classification systems. The study used various methods including emotional classification and clustering algorithms.

In general, previous work on understanding the mental health of students has investigated factors such as gender and academic year. The Center for Addiction and Mental Health [6] found through surveys that females were more prone to report mental health issues than males, and final-year students were least likely to report these symptoms as compared to students in other years. The National Union of Students [5] conducted a survey of colleges and universities across Scotland found that examinations, concerns about future career prospects and finances were the major sources of stress. Zacaj [9] studied gender differences in

Engineering education, focusing on women's integration in the classroom, outlook for future careers, decision-making power and responsibility. Soet and Sevig [7] investigated the effect of ethnicity and sexual orientation on various mental health problems (depression, eating disorders, substance use, etc.). One insight from this study was that African American students were found to be less distressed than their counterparts. Finally, Trockel et al. [8] studied the academic performance of first-year college students and their health behaviour. Their results demonstrate that feelings of anxiety, depression, and time pressure negatively affected the performance of these students. However, participating in extracurricular activities alongside having a good support system positively affected the academic performance. In this paper, we use regression and classification to reach similar conclusions to those reported in previous work (including finding that women and first-year students in Engineering tend to have lower mental health and that the number of hours of homework is strongly correlated with mental health), and we present new insights into the mental health of Engineering students enabled by the use of regression and classification algorithms.

3. METHODOLOGY

The first part of our survey included questions about potential academic influences on mental health, summarized in Table 1. Previous work has focused on factors such as financial situation and career prospects; in this study, we focus mainly on academic factors. In particular, we hypothesize that students enrolled in competitive programs with a high workload and an imbalance of extra-curricular activities will have lower mental health scores.

Table 1: Attributes used in the first part of the survey

Attribute	Possible Values
Gender	Male, Female
Year of study	1,2,3,4
Engineering program	Environmental, Electrical, Computer, Systems Design, Mechanical, Chemical, Geological, Civil
Hours of class per week	0-40
Hours of homework per week	0-80
In a committed relationship of more than 6 months?	Yes, No
Recent relationship break-up?	Yes, No
Hours of extracurricular activities per week (sports, student politics, student clubs, etc.)	0-40

The second part asked the participants to rate, on a scale from zero to six, the following five aspects of their mental health.

1. Ability to Enjoy Life: characterized by enjoying the present and worrying less about the future or the past.

2. Resilience: ability to recover from adversity; a characteristic shared by those who cope well with stress and change.

3. Balance in various aspects of life, such as time spent alone or with others, work and play, sleep and wakefulness, and rest and exercise.

4. Emotional Flexibility: ability to reduce stress that is obtained from rigid emotional expectations.

5. Self-Actualization: ability to recognize one's abilities and the process of this recognition.

These five aspects were defined by the Canadian Mental Health Association [4]. In addition to considering each aspect separately, we also summed up the five scores to compute an overall mental health score for each respondent (on a scale from zero to 30).

We conducted the survey online and on campus during a seven-day period in the Fall semester, targeting undergraduate Engineering students from all programs. We received 312 responses, which corresponds to 5.6 percent of the Engineering student population, and discarded six due to missing responses and/or values out of bounds. 70 percent of the respondents were male. 71 percent were single.

We used the WEKA data mining toolkit [2] to analyze the survey results. First, we applied least-squares linear regression to predict the overall mental health score as well as the individual five component scores. We then computed separate linear regression models for selected Engineering programs (we ignored programs from which we received very few responses), and separate models for each year of study (across all programs), to see if certain programs or years face unique mental health challenges.

Next, we discretized the numeric attributes. Mental health scores for each category were converted into: Very Low (0), Low (1), Medium (2-3), High (4), Very High (5) and Excellent (6). Overall mental health scores were converted into: Very Low (0-5), Low (6-10), Medium (11-15), High (16-20), Very High (21-25) and Excellent (26-30). Hours of class and extracurricular activities per week were converted as follows: Low (0-15), Medium (15-30), High (30-40); hours of homework per week were converted similarly. Using the transformed data, we computed prediction rules for overall mental health and for the five components of mental health using the PRISM algorithm. Then, as before, we separately computed rules for each program of study (across all years) and each academic year (across all programs).

Having computed the above regression and classification models, we examined the regression coefficients and classification rules to understand which attributes have the greatest effect on the mental health of students.

Finally, we conducted exit interviews with a sample of the respondents to help explain the results of our analysis.

4. RESULTS

According to our regression and classification results, the strongest signals in the data were as follows:

- In terms of overall mental health, the year of study and number of hours of homework had the greatest effect.
- Second-year students had the highest overall scores and first-year students had the lowest scores. Fourth-year students also had relatively low scores.
- In terms of academic programs, Electrical Engineering students had lower mental health scores due to the competitive nature of the program, while Systems Design students had higher scores due to strong classmate relationships and a flexible curriculum.
- Women in all Engineering programs have lower overall mental health (especially in Mechanical Engineering), but higher Emotional Flexibility.

- Self-Actualization was negatively affected by a high number of hours of class per week, but positively affected by a high number of homework hours per week.
- Being in a relationship reduces the Balance and Emotional Flexibility mental health components.

Table 2 summarizes the Root Mean Squared Error (RMSE) of the regression models and the prediction accuracy (the number of correctly classified instances divided by the total number of instances) of the PRISM classification rules for each mental health component. We used 10-fold cross-validation in each case. It does not appear that any one component is significantly easier or harder to predict using the features collected in our survey.

Table 2: RMSE and prediction accuracy for each mental health component

Component	RMSE	Prediction Accuracy
Ability to Enjoy life	1.29	80%
Resilience	1.32	84%
Balance	1.31	80%
Self-Actualization	1.53	83%
Flexibility	1.12	75%

In the remainder of this section, we discuss the above findings in more detail, paying particular attention to the impact of the academic program, year of study and gender on the overall mental health and its five components.

4.1 Impact of Year of Study

For each academic program within the Faculty of Engineering, we computed a regression model that predicts the overall mental health score based on four indicator variables corresponding to the four academic years. Table 3 shows the regression coefficients. Based on their magnitudes and signs, it appears that being in first year or in fourth year has a negative effect on the overall mental health score, with first year having the most negative impact. Exit interviews were conducted to determine the lead indicators of these results.

Table 3: Regression coefficients for predicting overall mental health for each academic program based on academic year

	First Year	Second Year	Third Year	Fourth Year
Electrical	-1.1848	0.9269	-0.0288	-0.1853
Environmental	-1.2735	0.922	-0.0748	-0.1589
Mechanical	-1.2312	0.9174	-0.0021	-0.1374
Civil	-1.235	0.9194	-0.0175	-0.1042
Systems Design	-1.159	0.8555	0.0714	-0.2464
Computer	-1.2241	0.9146	0.0085	-0.152
Geological	-1.1696	0.9022	-0.0725	-0.0732
Chemical	-1.2026	0.915	-0.0248	-0.13

Based on the exit interviews, first-year students found it difficult to be separated from family and friends from home. They also found that moving to a new city required lifestyle and workload adjustments. Furthermore, first-year Engineering students have a high classroom workload (35 to 40 hours of class per week), with many courses requiring weekly homework assignments, which adds to the high workload.

Second-year students were more relaxed due to the decreased focus on weekly evaluations. In particular, second-year students reported the lowest number of hours of homework. These students also found that the reduced focus on the fear of failure was a large stress reducer. As they had been successful passing first year, their confidence levels were increased.

However, by the time students reach the last year of study, many students find themselves overwhelmed with complex and lengthy course projects. Student debts are typically increasing and this compounds their worries. Late in the third year and early in the final year, some students feel that they are ready to finish their undergraduate career.

4.2 Impact of Academic Program

Both regression and classification results revealed that students from Systems Design Engineering tend to have higher scores. For example, the following rule was discovered by PRISM:

> If Systems Design Engineering = Yes and Year = Third then Overall Mental Health = Very High

Exit interviews suggested that Systems Design Engineering students expressed a high cohesion among classmates across all years. From the first day, these students are encouraged to make strong relationships. Students in this program found that their ability to be hired in any industry for internships proves their program to be fulfilling and accomplishing in real-world applications. Finally, Systems Design students have the freedom to focus on areas of interest through a more lenient course elective program.

On the other hand, being in Electrical Engineering was negatively correlated with mental health according to our regression and classification results. As per the exit interviews, the ultra-competitive nature of this program and the large number of compulsory courses (leaving little time for elective courses) appeared to be the main culprits.

4.3 Impact of Gender

In four of the five categories (Ability to Enjoy Life, Resilience, Balance, and Self-Actualization), being female negatively impacted the regression model for mental health across all programs and years. Being female was a positive factor only for Emotional Flexibility. Running PRISM to determine the rules associated with the mental health of women in engineering provided some specific insight into the Mechanical Engineering program as well as females in engineering as a whole:

> If Mechanical Engineering = Yes, and Gender = Female then Overall Mental Health = Low

> If Gender = Female then Flexibility = Very High

To gather some insight into these findings, exit interviews were conducted with female students. Being the minority in most classes, especially in Mechanical Engineering where the gender bias is severe, makes most women feel like they need to prove themselves to their professors, their male counterparts and in male-dominated industries on work terms. Most of the professors are male and the delivery of material is tailored towards men (humour during class is often directed towards male students). Lastly, there is a lack of female companionship and there are high

levels of competitiveness within the women in classes, which contributes to feelings of isolation and loneliness.

4.4 Factors Affecting the Five Components of Mental Health

Overall, Engineering students rated their ability to enjoy life more highly than other mental health aspects. According to our exit interviews, students recognized that "some things can't be changed", and when asked why, the responses ranged from technical to organizational and even philosophical barriers.

According to the linear regression results, being in the Electrical Engineering program had negative coefficients for every mental health category except Self-Actualization. Also, for students in Environmental Engineering, Self-Actualization was lower in the first and second year and higher in later years.

The number of hours of class per week had a negative effect on Self-Actualization across all programs, and a negative effect on the Ability to Enjoy Life for most programs. However, the number of hours of class per week had a positive coefficient for the other three categories (Emotional Flexibility, Balance and Resilience). On the other hand, the number of hours of *homework* had a positive effect on Self-Actualization, but a negative effect on every other category. Time in the classroom is mandated and there is no emphasis on discovery, but the time spent applying this knowledge on homework is self-motivated, which could aid in realizing one's full potential.

According to our regression results across all programs, the number of hours spent on extracurricular activities had a very small negative effect on the Flexibility category of mental health, and a small positive effect on all other categories.

We found that being in a committed relationship had a negative effect on Balance and Emotional Flexibility, most likely due to the difficulty of managing a demanding academic program and personal relationships. In particular, we obtained the following rule:

> If In a Committed Relationship of More than 6 Months = Yes and Year = First and Gender = Male then Balance = Low

Finally, according to linear regression results, having gone through a recent breakup yields a negative effect on the Ability to Enjoy Life and Emotional Flexibility.

5. CONCLUSIONS, RECOMMENDATIONS AND FUTURE WORK

In this paper, we presented a case study of how data mining may be used to understand factors affecting the mental health of students. We applied linear regression and classification algorithms to mental health surveys completed by Engineering undergraduate students, which revealed interesting relationships between various aspects of mental health and the academic program, year of study, gender, workload and relationship status.

The results of this study suggest a number of recommendations to help improve the mental health of Engineering undergraduate students. Given that the number of hours of homework was an important factor, it may be beneficial to offer first-year students additional time-management training. Furthermore, more support should be provided to female Engineering students, e.g., counseling services or forums to invite women to talk about their experiences.

Our analysis of the impact of the year of study on mental health (recall Table 2) suggests that first-year Engineering students are under pressure to succeed and more support should be provided to them. Furthermore, it appears that fourth-year students experience the pressure of multiple course projects. One solution may be to spread out project-heavy courses through the second and third year rather than leaving all of them till the last year.

Another interesting finding was that students in a highly competitive and challenging program (Electrical Engineering) tend to have low overall mental health scores but high Self-Actualization scores, whereas System Design Engineering students have high overall scores, which are possibly due to the flexible curriculum. Further research should be done on balancing the number of required and elective courses while maintaining a challenging and practical curriculum.

In this paper, we focused mainly on academic factors that may affect mental health, such as year of study, program and workload. An interesting direction for future work is how to reliably collect and analyze (in a manner that ensures data privacy) data describing other factors that may affect the mental health of students, such as instructor-class relationships, grades, and satisfaction with work placements in co-operative education.

6. REFERENCES

[1] Diederich, J., Al-Ajmi, A., Yellowlees, P., (2007). *Ex-ray: Data Mining and Mental Health*, Applied Soft Computing, 7(3):923-928.

[2] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H., (2009). *The WEKA Data Mining Software: An Update*, SIGKDD Explorations, 11(1):10-18.

[3] Li, H., Li, W., Liu, Q., Zhao, A., Prevatt, F., Yang, J., (2008). *Variables Predicting the Mental Health Status of Chinese College Students*, Asian Journal of Psychiatry, 1(2):37-41.

[4] Meaning of Mental Health, Canadian Mental Health Association (2013), retrived from http://www.cmha.ca/mental_health/meaning-of-mental-health/#.UHr5GVH08YQ.

[5] National Union of Students Scottland, (2010). *Silently Stressed: A Survey into Student Mental Wellbeing*, retrieved from http://www.nus.org.uk/PageFiles/12238/THINK-POS-REPORT-Final.pdf.

[6] Paglia-Boak, A., Adlaf, E.M., Hamilton H.A., Beitchman, J.H., Wolfe, D., Mann R.E., (2012). *The Mental Health and Well-Being of Ontario Students, 1991-2011: Detailed OSDUHS Findings* (CAMH Research Document Series No. 34). Toronto, ON: Centre for Addiction and Mental Health.

[7] Soet, J., Sevig, T., (2006). *Mental Health Issues Facing a Diverse Sample of College Students:Results from the College Student Mental Health Survey*, Journal of Student Affairs Research and Practice, 43(3):786-807.

[8] Trockel, M.T., Barnes, M.D., Egget, D.L., (2000). *Health-Related Variables and Academic Performance Among First-Year College Students: Implications for Sleep and Other Behaviors,* The Journal of American College Health, 49(3):125-131

[9] Zacaj, A., (2010). *Gender Differences in Engineering Education*, M.A.Sc. Thesis, University of Waterlo