**Natural Language Processing in Engineering Education**

Neha Kardam

Department of Electrical and Computer Engineering

University of Washington

Seattle, USA.

*Abstract-* Natural language processing (NLP) techniques are used in this research project to investigate peer support and faculty support students reported in over 1000 surveys. Students from 19 junior and sophomore level electrical and mechanical engineering courses at a large public research university reported existing and preferred levels of peer support and faculty support in in-person and remote learning contexts. The survey data consists of one short answer question for peer support and one for faculty support. In the case of peer support, short answer questions were deductively coded using a cooperative learning framework, whereas responses in faculty support were not coded. In peer support, I seek to solve the multi-label classification problem by predicting four peer support categories. In the case of faculty support, I broke down the problem into two parts. First, to identify the topics that have emerged from the students' responses, and then to conduct sentiment analysis on the students' responses based on demographics.

The findings of this study will aid in identifying the obstacles that students have encountered in terms of the peer and faculty support they require in engineering education, allowing them to receive any necessary assistance in having a pleasant learning experience.

1. **INTRODUCTION**

Peer and faculty support are critical components of promoting student involvement and academic achievement. Both play a key role in shaping students' learning and providing the best possible college experience.

*Peer support*

For many college students, the undergraduate years are the first time that they have spent significant time away from their families. Not surprisingly, students begin to spend most of their social time with their peers and draw much of their social and emotional support from their peer groups [1]. The existing knowledge base repeatedly validates the importance of peer support in both social and academic systems in college. Studies of engineering undergraduates have demonstrated that the degree to which students perceive respect from peers in their classes is positively and significantly correlated to satisfaction. Students are better able to cope in college

with peer support [2] [3], which in turn results in improved academic outcomes [4].

*Faculty Support*

Support from faculty is an important element of student engagement in higher education institutions. Existing literature demonstrates the important role that faculty interactions, both formal and informal, can play regarding students’ social and academic integration—known elements of students’ successful degree completion [5]. Social engagement and feelings of connectedness at higher education institutions are heavily influenced by faculty support. Faculty support also has the potential to exceed the influence of overall student culture with regard to students’ social engagement [6]. The care and respect that faculty members express has been associated with positive experiences and an emotional commitment to an institution [7].

Students, whether in a physical or virtual learning environment, should ideally have access to all of the resources they require to communicate with peers and instructors. However, we feel that there may be some gaps in meeting those needs with rapidly evolving teaching methodologies, whether in-person or online. This study will forecast what current students will experience in terms of peer support and faculty support. The findings will help in providing resources to academicians in order to address those needs and deliver the best learning experience possible in the future.

1. **PROBLEM DESCRIPTION**

The aim of this project is to solve two problems as mentioned below.

Problem #1: Peer Support

The goal is to categorize/classify student responses to the following question:

*"What one action can students in your lab groups take to improve your educational experience at a large public research institution?"*

This study employs the cooperative learning framework developed by Johnson and Johnson [8] [9]. Cooperative learning is defined by researchers as a method that can be applied to a variety of situations, such as teaching

specific content (formal cooperative learning groups), ensuring effective cognitive review of information during a lecture or presentation (informal cooperative learning groups), and providing long-term support for academic achievement (cooperative base groups) [10]. The cooperative model is described by five key categories:

Category 1: Individual Accountability (IA)

Category 2: Positive Interdependence (POSI)

Category 3: Promotive Interactions (PI)

Category 4: Interpersonal and Social Skills (ISS)

Category 5: Group Processing (GP)

Note: no student responses were classified as Category 5.

Two research assistants coded the short answer responses from all the students in the survey population using the cooperative learning framework previously described.

The purpose of this section of the NLP project is to classify student responses into the four categories above for maximum accuracy, where accuracy is defined as a classification that is identical to what the researcher determined in initial coding and analysis.

Problem #2: Faculty Support

The goal is to identify the best number of clusters that responses to the following question organize into using NLP methods:

*"What one action can faculty take to improve your educational experience at a large public research institution?"*

No assumptions are made about how many clusters (groups) these responses will fall into. The goal of this portion of the NLP project is to identify the optimal number of clusters to support future coding of these responses. This will be accomplished by applying 3 NLP techniques to represent the students' responses as follows:

1) Topic Modelling

2) Sentiment analysis

1. **DATA INFORMATION**

This study is part of a larger, single-institution research project that evaluated the connections between various forms of support (from faculty, TAs, and peers) and multiple forms of course-level engagement (attention, participation, effort, positive and negative emotional engagement) both in person and in remote learning settings.

*Participants*

The study includes a student population of 1328 undergraduate students recruited across four engineering majors and 19 separate classes at the sophomore and junior levels in both traditional and remote learning settings from a large public research institution located in an urban area. Self-reported ethnicity was Asian (43.3%), Black (2.9%), Hispanic (3.1%), White (40.7%), Pacific-Islander (less than 1%), Native American (less than 1%), and Other (2.6%).  A number of students were mixed race (6.7%) among which White-Asian was most common (73%). Approximately 25.7% of the original sample was female, with 73.8% male and less than 1% reporting as non-binary. Students also reported their status as U.S. citizens (78.4%), Permanent Residents (5.1%), or International (16.3%) with the most common countries of origin being China (16.2%) or India (4.2%). Preliminary data analyses for all the scales indicated no emerging differences between permanent residents and citizens, so these groups were combined in the final analyses, accounting for 83.5% of the sample population. Similarly, all international students were combined into a single group (16.3% of the sample population) as supported by the preliminary analyses of all the measurement scales. The final demographic characteristics of the student population are summarized in Table 1.

12 courses were surveyed during remote learning (post pandemic) in Spring of 2020 and 7 courses

were surveyed during in-person learning between 2016 and 2018.  As a result, 45.7% of survey respondents reported their experiences in remote learning while 54.3% reported their experiences in traditional college classrooms.  Of the 19 courses studied, 7 were taught in both in-person and remote learning and three (ME1, ME2, EE4) were taught by the same instructor in both settings.

*Procedures*

IRB (Internal Review Board) approval (STUDY00000378) was obtained to recruit and survey undergraduate students for this study. Researchers interacted with faculty teaching the courses relevant to this study, but the researchers did not engage directly with the students. Since the vast majority of the students in this study began their academic careers at four-year colleges, it was unlikely that they had any direct experience of online learning prior to spring 2020. All participation was voluntary in this study, and students were informed that their survey responses would remain confidential. In several courses, students were incentivized with a nominal amount of extra credit for the course in which they were recruited. In one traditional learning course (EE1), students completed a paper-and-pencil copy of the survey, while in all remaining courses, students completed an electronic survey online and outside of class. Some students were present in more than one class; since survey questions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 1: Courses Studied** | | | | | |
| **Course** | **Level** | **Setting** | **Topic** | **Enrolled** | **Participants (*N*)** |
| ME1 | Sophomore  Junior  Senior | Traditional | Visualization & CAD | 179 | 140 |
| Remote | 155 | 73 |
| ME2 | Sophomore | Traditional | Engineering Statics | 69 | 20 |
| Remote | 92 | 6 |
| ME3 | Sophomore | Traditional | Kinematics and Dynamics | 263 | 218 |
| Remote | 184 | 143 |
| EE1 | Sophomore | Traditional | Introduction to Electrical Engineering | 223 | 175 |
| Remote | 105 | 73 |
| EE2 | Sophomore | Traditional | Circuit Theory | 91 | 70 |
| Remote | 69 | 57 |
| EE3 | Sophomore | Traditional | Continuous Time Linear Systems | 84 | 63 |
| Remote | 86 | 70 |
| EE4 | Sophomore | Traditional | Digital Circuits and Systems | 41 | 35 |
| Remote | 37 | 27 |
| EE5 | Junior | Remote | Devices and Circuits I | 56 | 49 |
| EE6 | Junior | Remote | Devices and Circuits II | 36 | 25 |
| EE7 | Junior | Remote | Discrete Time Linear Systems | 47 | 37 |
| EE8 | Junior | Remote | Energy Systems | 71 | 37 |
| EE9 | Junior | Remote | Applied Electromagnetics | 15 | 10 |

referred to a specific class ("this class"), duplicate surveys were retained for analysis. All results were cross-sectional.

1. **METHOD USED AND EXPERIMENTAL RESULTS**

This section of the research study will describe the technique used to tackle problems #1 and #2, as well as the results. Jupyter Notebook was used to implement the model for this project using python language.

**Problem #1: Peer Support**

This is a multi-label classification problem that requires two critical pieces of information to predict the labels associated with students’ comments. The first piece of information required is the students' responses to the peer support question, followed by the labels (PI, IA, ISS, POSI) associated with the students' responses. The frequency of responses in each category is summarized in Table 2.

In the analysis, we removed the responses that were categorized under other, off-topic and none. The data is further preprocessed by removing all the students’ responses where any row contain a null value or empty string. Next step was to convert the string values of the labels into numerical values as shown below in the Table 3.

A function is created to clean the text data by eliminating punctuation, numbers, and spaces, as well as single characters. We then divided our data into training (80%) and test (20%) sets. To transform text input to numeric equivalents, the text is turned into embedded vectors using Glove word embeddings (6B tokens, 400k vocab, 50d).

|  |  |  |
| --- | --- | --- |
| **Table 2: Student Expectations of Peer Support** | | |
| **Primary Category** | **Frequency of Response (%)** | |
| **In-person** | **Remote** |
| Other | 14 (2.93%) | 7(1.10%) |
| Off-Topic | 60(12.6%) | 77(12.2%) |
| None | 48(10.0%) | 60(9.50%) |
| Promotive Interaction (PI) | 130(27.3%) | 216(34.2%) |
| Individual Accountability (IA) | 135(28.3%) | 105(16.6%) |
| Interpersonal/Social Skills (ISS) | 52(10.9%) | 76(12.4%) |
| Positive Interdependence (POSI) | 38(7.96%) | 93(14.7%) |
| **Total** | **477** | **631** |

Furthermore, a multi-label classification model is created in a neural network using the dense layer, which is a deep-connected neural network layer, meaning is that each neuron in the dense layer gets input from all neurons in the preceding layer. For problem #1, we used two dense layer output models to predict the labels as mentioned below:

*1)Multi-label text classification with single dense output layer Model*: This is our baseline model which has a single dense layer with four outputs with sigmoid activation and binary cross entropy loss functions. In the output dense layer, each neuron will represent one of the four output labels. For each neuron, the sigmoid activation function will yield a value between 0 and 1. If the output value of any neuron is larger than 0.5, it is presumed that the comment belongs to the class represented by that neuron [11].

|  |
| --- |
| **Table 3: Numerical labeled data** |
| A screenshot of a phone  Description automatically generated with low confidence |

The model script used creates one input layer, one embedding layer, one LSTM layer with 128 neurons and one output layer with 4 neurons since we have 4 labels in the output. I trained the model with 5 epochs and with adam as an optimizer after that the model is evaluated on the test set. Model summary has been given below in table 4.

|  |  |  |
| --- | --- | --- |
| **Table 4: Single dense output layer model summary** | | |
| ***Layer (type)*** | ***Output Shape*** | ***Param #*** |
| input\_3 (Input Layer) | [(None, 200)] | 0 |
| embedding\_2 (Embedding) | (None, 200, 50) | 72100 |
| lstm\_2 (LSTM) | (None, 128) | 91648 |
| dense\_5 (Dense) | (None, 4) | 516 |
| Total params: 164,264  Trainable params: 92,164  Non-trainable params: 72,100 | | |

Our model achieved an accuracy of 47%. Finally, the plot of accuracy values for training and test sets is created to see if our model is overfitting.

*2) Multi-label text classification with multiple dense output layer model*: In the multiple dense output layer, for each label, one dense output layer is created. The output will contain a total of four dense layers. The sigmoid function will be different for each layer [11].

The data is cleaned and divided in this model in the same way as it was done in the previous model. In this model, however, we want to build a separate output layer for each label. As a result, we generate four variables to store individual labels from the data. The text data input is then converted into the embedded vectors using glove word embedding (6B tokens, 400k vocab, 50d). The model architecture is built with one input layer, one embedding layer, and one LSTM layer of 128 neurons. The LSTM layer’s output will be utilized as the input to the four dense output layers. One neuron with sigmoid activation function will be present in each output layer. For the appropriate label, each output will predict an integer value between 1 and 0. Model summary has been given below in table 5.

After assessing the model’s performance, an accuracy of 69% is achieved on the test set via multiple output layers.

**Problem #2: Faculty Support**

The faculty support problem is broken into two parts: A) Topic modeling, B) Sentiment Analysis. Each approach employed has been explained in detail below, along with the findings.

**A) Topic Modelling:** Topic modeling is an unsupervised methodology for analyzing vast amounts of text data by grouping texts together. The text data in topic modeling does not have any labels linked to it. Topic modeling, group the documents into clusters based on common characteristics. In this research study we have used two approaches for topic modeling: Latent Dirichlet Allocation and Non-Negative Matrix factorization.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 5: Multiple dense output layer model summary** | | | |
| ***Layer (type)*** | ***Output Shape*** | ***Param #*** | ***Connected to*** |
| input\_2 (Input Layer) | [(None, 200)] | 0 |  |
| embedding\_1 (Embedding) | (None, 200, 50) | 72100 | input\_2[0][0] |
| lstm\_1 (LSTM) | (None, 128) | 91648 | embedding\_1[0][0] |
| dense\_1 (Dense) | (None, 1) | 129 | lstm\_1[0][0] |
| dense\_2 (Dense) | (None, 1) | 129 | lstm\_1[0][0] |
| dense\_3 (Dense) | (None, 1) | 129 | lstm\_1[0][0] |
| dense\_4 (Dense) | (None, 1) | 129 | lstm\_1[0][0] |
| Total params: 164,264  Trainable params: 92,164  Non-trainable params: 72,100 | | |  |

*1)Latent Dirichlet Allocation (LDA):* It is a generative statistical model that allows unobserved groups to explain why some sections of the data are similar. Mathematically, in LDA, documents represent a probability distribution over latent topics, whereas topics represent a probability distribution over words [12].

Before applying the LDA, we first pre-process the data and use the count vectorizer to construct a vocabulary of all the terms in our data. We specify that only terms that appear in less than 80% of the responses and in at least two responses are included. We also delete all stop words because they contribute nothing to topic modeling. Then, for each topic, we have used LDA to generate topics as well as the probability distribution for each word in our vocabulary.

To pick the 10 words with the highest likelihood for the first topic, we use the components\_ attribute and a 0 index as the value. The argsort() method is used to sort the indexes according to their probability values. Following sorting, the 10 words with the highest probability are given to the array’s last 10 indexes. Then we retrieve the words with the highest probability. After that, we discovered ten words with the highest probability for each of the three topics as shown in table 6. Finally, the probability of each topic will be assigned to each student’s response using LDA.transform() function.

*2) Non-negative matrix factorization (NMF):* Non-negative matrix factorization is a supervised learning approach that also conducts clustering and dimensionality reduction. To do topic modeling, it may be utilized in conjunction with the TF-IDF (term frequency-inverse document frequency) method [12].

|  |  |
| --- | --- |
| **Table 6: Topics emerged from Latent Dirichlet Allocation** | |
| ***Topics*** | ***Top 10 words*** |
| #1 | [‘like’, ‘material’, ‘make’, ‘classes’, ‘professors’, ‘questions’, ‘time’, ‘lectures’, ‘class’, ‘students’] |
| #2 | [‘exam’, ‘provide’, ‘extra’, ‘help’, ‘exams’, ‘problems’, ‘practice’, ‘office’, ‘hours’, ‘nan’] |
| #3 | [‘material’, ‘professor’, ‘online’, ‘helpful’, ‘clear’, ‘slides’, ‘class’, ‘lectures’, ‘notes’, ‘lecture’] |

In NMF, the document term matrix is created using TF-IDF, and then the probability matrix is constructed, which comprises the probabilities of all the words in the lexicon for all topics. The next stages in the NMF are similar to those in the LDA in that we identify the top ten words, and then we discover the top ten words with the highest probability for the top five topics. Table 6 shows topics that are similar to those obtained through the NMF. After reviewing all of the topics from both techniques, three themes seemed to have emerged: Assessment, Supporting Materials, and Faculty Interaction

**B) Sentiment Analysis**: In this research study, sentiment analysis is performed to determine the

students’ feelings towards faculty support. Pretrained sentiment analysis tools are used to do sentiment analysis, as listed below:

1) NLTK: The Vader sentiment analysis tool from NLTK uses a bag of words technique (a lookup table of positive and negative phrases) in conjunction with some simple heuristics [13].

2) Textblob: Sentiment Analysis in textblob uses a bag of words classifier in the same manner as NLTK does, but it also incorporates Subjectivity Analysis [13].

3) Flair: Flair’s sentiment classifier is built on a character-level LSTM neural network that predicts based on letter and word sequences [13].

Table 7 displays the results of all three pre-trained sentiment analyses. In the table, SA1 column represents the student’s responses to faculty support, and the columns to the left of SA1 indicate the student's information, such as social class, father's education, gender, age, major, and so on. Using NLTK,we discovered negative, neutral, positive, and compound parameters on the right side of SA1. Textblob provided polarity and subjectivity, while flair provided positive/negative sentiments.

Figure 1 illustrates the gender differences in student sentiments toward faculty support.

|  |
| --- |
| **Table 7: Sentiment Analysis for faculty support** |

***Graphical user interface

Description automatically generated***

Figure 1 illustrates the gender differences in student sentiments toward faculty support. We observed that female students had around 2% higher positive faculty support experience than male students.

Table 7 shows students' sentiments towards faculty support by demographics. We discovered that white students and Asian students received more than 70% of faculty support, whereas

|  |  |
| --- | --- |
| Chart, pie chart  Description automatically generated | Chart, pie chart  Description automatically generated |
| **Fig 1: Sentiment Analysis by gender (Faculty support)** | |

Pacific Islander and black students received more than 40% of negative faculty support.

|  |  |  |
| --- | --- | --- |
| **Table 7: Sentiment Analysis by demographic (faculty support)** | | |
| **Students Demographic** | **Positive faculty support** | **Negative faculty support** |
| White | 75.9% | 24.1% |
| Asian | 70.1% | 29.9% |
| Black | 59.3% | 40.7% |
| Hispanic | 67.6% | 32.4% |
| Pacific Islander | 50% | 50% |
| Other | 67.7% | 32.3% |

1. **CONCLUSION**

This paper summarizes the findings from NLP applications in engineering education. This study predicted different labels of peer support based on student survey responses. Deep neural network approaches were used to achieve a maximum accuracy of 69 percent in multi-label classification. Using LDA and NFM techniques, we observed that the topics - assessment, supporting material, and faculty interaction evolved from students' responses with regards to faculty support. Additionally, we revealed students' attitudes regarding faculty support by gender and demography using pre-trained sentiment analysis methods. There were no significant differences in positive and negative emotions across genders, although there were notable differences in perceived faculty support across demographic groups. This study may be expanded in the future to see whether there are any biases in the findings obtained for multi-label classification when the labels were coded deductively. Additionally, the repercussions of faculty support may be examined in detail in order to ascertain the underlying reasons for students receiving varied degrees of perceived support from instructors, as well as ways for closing those gaps. Semantic similarity may also be used and investigated in detail to discover groups of students with comparable faculty support depending on their demographics.

1. **LIMITATION**

This study focused on the student experience at a single institution and in a limited number of majors (primarily electrical and mechanical engineering), therefore the generalizability to other academic settings may be limited. Despite the fact that no survey questions asked if students had prior online survey experience, the vast majority of students in this study began their college education at the traditional, in-person institution participating in this study. As a result, it is fair to say that most students have little or no exposure to online classes prior to Spring 2020. Further, the remote learning data were collected in the first full term of remote learning and do not reflect longer term adjustments that students may have made as remote learning extended into the 2020-2021 academic year. Another drawback is that the data obtained from the courses in this study was taught by different professors, and the survey was conducted outside of the control of the researchers. The qualitative analysis of responses to short answer questions provides important insight as to what kinds of behaviors and contributions students expect from their peers and what they believe to be the most important contributions from their peers and faculties. Therefore, despite the limitations, the results of this study offer a rich perspective regarding what students perceive as necessary support from their peers and instructors.

**REFERENCES**

[1]  A. Astin, “What matters in college? : Four critical years revisited,” 1st ed., Jossey-Bass higher and adult education series, San Francisco: Jossey-Bass, 1993, 15, pp. 54.

[2] C.T. Amelink, and E.G. Creamer, “Gender differences in elements of the undergraduate experience that influence satisfaction with the engineering major and the intent to pursue engineering as a career,“ *Journal of Engineering Education*, vol. *99*(1), pp.81-92,2010.

[3] P.T. Terenzini, L.I. Rendon, M.L. Upcraft, S.B. Millar, K.W. Allison, P.L. Gregg, and R. Jalomo, “The transition to college: Diverse students, diverse stories,” *Research in higher education*, vol. *35*(1), pp.57-73, 1994.

[4]  S. Ramsay, E. Jones, and M. Barker, “Relationship between adjustment and support types: Young and mature-aged local and international first year university students,” *Higher education*, vol. *54*(2), pp.247-265, 2007.

[5] V. Tinto, *Leaving College: Rethinking the Causes and Cures of Student Attrition. Second Edition*. Chicago, IL: University of Chicago Press, 1993.

[6] M. A. Lamport, “Student-faculty informal interaction and the effect on college student outcomes: a review of the literature,” *Adolescence*, vol. 28, no. 112, pp. 971-990, 1993.

[7] V. LaMastro, “Influence of perceived institutional and faculty support on college students’ attitudes and behavioral intentions”, *Psychological Reports*, vol. 88 no. 2, pp. 567-580, 2001, doi: 10. 2466/pr0.2001.88.2.567.

[8] D. W. Johnson, R. T. Johnson and K. A. Smith, “Cooperative learning returns to college what evidence is there that it works?,” *Change: the magazine of higher learning*, vol. *30*(4), pp.26-35, 1998.

[9] De Winter, J. F. C., & Dodou, D. (2010). “Five-point likert items: t test versus Mann-Whitney-Wilcoxon”, *Practical Assessment, Research, and Evaluation*, vol. *15*(1), p. 11, (Addendum added October 2012).

[10] David W. Johnson, Roger T. Johnson. “Making cooperative learning work”, Theory into practice 38, no. 2, 67-73, 1999.