Analysis of the impact of COVID-19 on student life based on the results of a survey

Takato Kamimura   
*Faculty of Engineering  
Graduate School of Engineering*  
*Nagoya Institute of Technology* Nagoya, Japan   
t.kamimura689@stn.nitech.ac.jpHirotaka Itoh  
*Information Technology Center   
Nagoya Institute of Technology*Nagoya, Japan   
ht-itoh@nitech.ac.jpKenji Funahashi  
*Information Technology Center*   
*Nagoya Institute of Technology*Nagoya, Japan  
kenji@nitech.ac.jp

*Abstract*— The spread of COVID-19 has adversely affected many sectors, including tourism, retail, and manufacturing. The educational field is no exception, and many universities, including our own, have taken measures to prevent infection, such as implementing online classes and banning the use of facilities. These infection control measures are expected to change the living environment of students. If students are unable to lead their lives as before due to changes in their living environment, this may lead to a decline in academic performance and poor health. Therefore, it is very important for universities to understand how COVID-19 affects students' lives in order for them to lead healthy student lives.

Therefore, this study aims to understand the impact of COVID-19 on students' lives by using data mining techniques to analyze the response data from a survey conducted for students, and to provide appropriate support and infection prevention measures for students.

As a result, we identified a tendency for students to feel anxious about infection with COVID-19 and changes in students' evaluations of classes conducted in a face-to-face format under infection prevention measures. We believe that these results can contribute to reconsideration of support for students and class formats.

Keywords—Data Mining, Support, COVID-19

# Introduction

In recent years, the rapid development of information and communication technology and hardware has led to the utilization of large-scale data known as big data. Data mining technology is necessary for the correct and effective use of big data and is used in various fields, including the medical, manufacturing, and retail industries. Examples include the automatic estimation of factors causing defects in a manufacturing line [1] and the analysis of text-type data of product coupon introductions [2] to promote customers' willingness to purchase. Data mining technology is also used in a wide range of other fields, such as in predicting when highway equipment will fail [3]. Similarly, in the field of education, a system that uses text mining to provide feedback from class evaluation questionnaires has been developed [4]. When data mining technology is used in education, the responses in questionnaires administered to students, such as class evaluation questionnaires, are often used as data [5].

Many sectors, including tourism, retail, and manufacturing, have been adversely affected by COVID-19, which spread worldwide and was characterized as a pandemic. The education sector was affected as well, and many universities implemented various measures to prevent infection, such as teaching classes online and banning the use of facilities. In addition, the lives of students were changed by these measures [6].

In this study, we used data mining techniques to analyze how students' lives have changed as a result of the COVID-19 pandemic and aimed to use the information obtained from the results to help provide appropriate support to students in the areas of study and finances.

Furthermore, we analyzed the tendencies of students who were anxious about COVID-19 and confirmed that they experienced various inconveniences in their school life.

# Method

## Data Used in the Analysis

The Nagoya Institute of Technology conducts an annual survey of students, called the Survey of Student Life. This study used the results of surveys conducted in 2020 and 2021. We collected responses to items regarding part-time job activities from the 2020 survey. Moreover, we used data from responses to questions regarding “implementation status,” “purpose,” “type of work,” “working hours/frequency,” “income,” “occupations in part-time jobs (first and second),” “purpose of conducting part-time work,” and “concerns about infection.” For those responses that could be quantified, positive responses were quantified as positive and negative ones as negative values, using 0 as the standard. We also used data from open-ended responses to the surveys conducted in 2020 and 2021.

## Correlation Analysis

Correlation analysis is an analytical method in which the strength of the relationship between two parameters is calculated and expressed as a numerical value called the correlation coefficient [7]. The correlation coefficient has a value between ‒1 and 1. The closer the absolute value of the correlation coefficient is to 1, the stronger is the relationship between the parameters. In this study, the concerns about infection, school year, and questionnaire items listed in Section II.A were used for correlation analysis.

## Decision Tree

Decision tree analysis is a method that divides data into explanatory and objective variables and classifies them according to classification criteria, such as whether the objective variable belongs to a category of explanatory variables or whether the explanatory variable is above or below a specific value. Decision tree analysis enables the visualization of the rules and factors of the objective variable. Although there are several algorithms for creating decision trees, this study uses an algorithm based on the degree of improvement in the Gini coefficients. In this study, the items listed in Section II.A, including “occupations in part-time jobs (first and second)” and “purpose of conducting part-time work,” were used to for decision tree analysis.”

## Random Forest

Random forest is a learning method called ensemble learning, in which multiple decision trees are created and class classification is performed by majority voting of the classification results of each tree. Since each decision tree can be processed in parallel, the computation can be performed at high speed. Random Forest trees are created using 80% of the data, and classification is performed on all data. By checking how accurate the classification is on the data used to create the tree, we can confirm how useful the explanatory variables are in classifying the objective variable.

## Co-occurrence Frequency Analysis

Two words that appear in the same sentence are called co-occurring words. Co-occurrence frequency analysis is a method of analyzing the frequency with which two words appear as co-occurring words to determine which topics are frequently discussed together. In this section, the results of co-occurrence frequency analysis are illustrated as a network diagram to analyze which words co-occur and their co-occurrence frequency.

# Results

## Correlation Analysis

Correlation analyses between the “concerns about infection” and school year and between the “concerns about infection” and various questionnaire items related to part-time work were performed. The correlation coefficients obtained are listed in TABLE I. Moreover, a *t*-test (e.g., *p* < 0.05) was used to determine the significance of each correlation coefficient. The items for which a significant correlation was found are indicated by “\*” next to the values ​​in the table.

TABLE I shows that there is a significant positive correlation only with the variable indicating work status before the spread of the infection. This suggests that the level of anxiety about infection tended to be slightly higher for those who worked more days per week before the spread of the infection.

The data treated in this study included a large number of first-year students who were already infested with COVID-19 at the time of admission, which may have influenced these results.Therefore, a correlation analysis by grade level was performed on the data used in TABLE I.

Correlation analysis by academic year from TABLE II also shows small correlation coefficients among most of the items. Weak positive correlations were found for "Average hours worked per week before infection spread" among the past year students and graduate students.

1. Correlation coefficients

| ***Questionnaire item*** | Correlation coefficient between concerns about infection and items |
| --- | --- |
| School year | 0.108 |
| Average hours worked per week before infection spread | ‒0.129\* |
| Average number of working days per week before infection spread | 0.051 |
| Average hours worked per week after infection spread | ‒0.147\* |
| Average number of working days per week after infection spread | ‒0.021 |
| Changes in working hours before and after the spread of infection | -0.065 |

1. Correlation coefficients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## Decision Tree

Decision tree analysis was conducted using the questionnaire items listed in Section II.C as explanatory variables. Fig. 1 shows Decision tree. To avoid overlearning, the depth was limited to 3 and the maximum number of branches to 5. In Fig. 1, leaves are shown as squares, and non-leaves as ovals. All variables described in Section II.C were used to create the tree except for “occupations in part-time jobs (first and second)” and “purpose of conducting part-time work.” As shown in Fig. 1, there was an association between work status before the spread of infection and high levels of infection anxiety, as indicated by the double-lined explanatory variables and leaves. This and the results of the correlation analysis suggest that part-time work status prior to the spread of COVID-19 may have affected students' subsequent feelings of anxiety about the infection.

However, the variables used for bifurcation were eliminatively selected from the variables prepared for bifurcation. variables that are not really effective in classifying the level of infection anxiety. It is possible that the variables used in the branching were eliminatively selected from the variables prepared for the branching and were not actually effective in classifying the concerns about infection.

Therefore, we checked to see how accurately the trees were classified. The correct response rate was 44.7%. Therefore, a random forest analysis was conducted to confirm the usefulness of the variables used.

TABLE III shows that the accuracy of the tree created is just over 50%. Although the system is better than a single decision tree, there is little increase in correctness with an increase in the number of parallel runs.The increase in the correctness rate by increasing the number of parallel runs is almost negligible.Therefore, it should be noted that the variables used to branch the tree may not be valid.

![ダイアグラム

自動的に生成された説明]()

1. Decision tree (excerpts)

TABLE Ⅲ. RANDOM FOREST

|  |  |
| --- | --- |
| **Number of trees** | **percentage of**  **correct answers** |
| 20 | 51.0 |
| 50 | 52.7 |
| 100 | 52.3 |

## Co-occurrence Frequency Analysis

To ensure that the number of co-occurrence relationships was equal, co-occurrence relationships were extracted for data with a co-occurrence frequency of 3 or higher for the questionnaire data from 2020 and for data with a co-occurrence frequency of 2 or higher for the questionnaire data from 2021.

Fig. 2 shows that the word “lecture” co-occurs with the positive words “pleasant” and “thankful” and the word “face-to face” which means a face-to-face lecture. This indicates that students' evaluations of face-to-face lectures are positive, although COVID-19 infection prevention measures have reduced the number of face-to-face lectures.

However, in Fig. 3, the negative word "anxiety" co-occurs with the word "face-to-face," which may refer to a face-to-face lecture. This indicates that over the course of the year, students' evaluations of face-to-face lectures shifted from positive to negative. The spread of COVID-19 has caused the University to move lectures online, but some lectures were still conducted in person. It is expected that students who initially had a positive impression of fewer face-to-face lectures changed their evaluation of face-to-face lectures after living in the COVID-19 disaster for a year.

This suggests that the form of lecture should be varied depending on the situation in order to keep students motivated to protest.

ダイアグラム

自動的に生成された説明

Fig.2 Collocational network using data from 2020(excerpts)

ダイアグラム

自動的に生成された説明

Fig3 Collocational network using data from 2021(excerpts)

# Conclusion

In this study, we analyzed the results of a survey on student life at the Nagoya Institute of Technology to examine the impact of COVID-19 on student life. We confirmed that COVID-19 affected student life, and for students who worked part-time, the frequency of work prior to the spread of infection might have been related to the increase in students' concerns about COVID-19. The co-occurrence frequency analysis showed that students' evaluations of “class” changed from positive to negative.

Future issues include reanalyzing the data with a larger data set, increasing the question format to broaden the data range of multiple-choice responses, and using student data from other universities. In addition, we would like to analyze the impact of COVID-19 on non-university institutions such as businesses.

We also believe that this study can contribute to improving future support and infection prevention measures for university students.

##### Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP19K03028.

##### References

1. H. Kondo, “Digital QC System Using Data Mining Technique for Production Line”, Toshiba review, vol. 58, No.7, 2003
2. J. Takakura, K. Igarashi, K. Suzuki, Y. Takahata, N. Fujisaki, Y. Matsui, H. Minami, “Catchphrase analysis of net sales sites with text mining”, Japanese Society of Computational Statistics, Vol.27, pp.97-98, 2013
3. K. Nishimura, M. Chiharu, M. Murayama, H. Yamaguchi, “Predictive Model Construction for Smart Highway Facilities Deterioration Using Data Minig”, IEEEJ Transactions on Industry Applications, Vol.140 No.10, pp.746-752, 2020
4. H. Matsuzaki, T. Saito, “Developmentand Evaluation of a Reporting System of the Student Course Evaluation UtilizingData and Text MiningTechnolog”, Japan Society for Educational Technology, Vol.35(3), pp.217-226, 2011
5. T. Naganawa, H. Itoh, K. Funahashi, “Analysis of the Relationship Between University Attendance and Grades by Data Mining”, Proc. IEEE- GCCE, pp.199-202, 2020.
6. T. Watanabe, “The impact of the COVID-19 on Impact on Consumption and Prices”, Monthly Capital Market, No.416, 2020. 4
7. K. Ishii, “Illustration: Understanding Data Mining”, The Nikkan Kogyo Shimbun, 2004