

Fundamentals of Data Science Final Report



*The Impact of Instructor Behaviour and Course Structure on Learning Effectiveness
in Higher Education*

Lecturer:

NUNUNG NURUL QOMARIYAH

Arranged By:

Joshua Efraim Rawatan 2702365000

Registan 2702364963

Abyan Ali Kartasasmita 2702355385

Fundamentals of Data Science

Computer Science Faculty

Binus International University

Table of Content

1. Introduction
 - 1.1 Genesis
 - 1.2 Problem Description
 - 1.3 Hypothesis
2. Related Works
 - 2.1 Key Framework
 - 2.2 Related Studies
 - 2.3 Research Gaps
3. Methodology
 - 3.1 Flowchart
 - 3.2 Dataset
 - 3.3 Preprocessing
 - 3.4 Models and Techniques
4. Discussion
 - 4.1 Result and Data Interpretation
 - 4.1.1 Linear Correlations
 - 4.1.2 Network Analysis
 - 4.1.3 Random Forest
 - 4.1.4 Performance Evaluation
 - 4.2 Statistical Analysis
 - 4.3 Strengths
 - 4.4 Weaknesses
5. Conclusion

Chapter 1. Introduction

1.1 Genesis

In higher education student's learning experiences are influenced by many factors, however, two of the most important factors are the instructor's behavior and the course structure. An instructor's approach to teaching, engaging, or any other method can make a great difference in how the students learn. Similarly, the way the course is designed from the pace and to the point of the course, plays an important role in how well the students will learn.

In every class, there are those who succeed due to the motivation they receive because of how engaging and approachable the instructors are or how clear and well-organized the course materials are. While others fail due to the course outline being confusing or the instructor did not teach the course well. With this observation, one question occurred to us: How do instructor behavior and course structure affect learning effectiveness in higher education, and what factors have strong and weak correlations?

Therefore, this report is designed to analyze the teaching methods and course design and their influence on learning effectiveness. By figuring out the behavior and course elements that will most effectively improve learning effectiveness, this report seeks to have a better understanding of the key factors that influence learning effectiveness in higher education.

1.2 Problem Description

In higher education, creating an environment that allows effective learning is a challenge that is influenced by many factors. While teaching methods and course structure are recognized as important, there are specific influences of the instructor's behavior and course structure on learning outcomes that are not fully understood. While the instructor's style of teaching and the structure of the course are an important factor affecting the learning experience, the exact nature of this relationship is not yet well understood. Due to these uncertainties, it makes it hard to identify which aspect of an instructor's behavior and course design are the closest linked with effective learning.

1.3 Hypothesis

As the course structure and instructor behaviour quality increases such as more engagement, clear definition of the courses, smooth teaching delivery, etc should show a positive correlation with learning effectiveness, influencing how the students retain the information and increasing the engagement of the students, meaning that there should be a higher attendance, decreased difficulty of the course, and less times that the course was repeated. A combination of supportive instructors and well-structured courses will result in a higher level of learning effectiveness compared to either factor alone.

Chapter 2. Related Works

2.1 Key Framework

Overall, the scope of this study is to investigate the impact of aspects in learning towards the learning and academic performance to find the underlying factor. This involves investigating the actions specifically the behaviour of the instructor and course structure on the student's academic performance. It also involves how the instructor creates assignments and the support that the instructor gives towards the students to help with learning.

2.2 Related Studies

1. Interpersonal Teacher Behaviour and Student Outcomes

This study investigates the interpersonal behaviours of teachers and their effect on students. It uses data of the responses of questionnaires of students to get the behaviour aspect of the study. This data was derived from 2 different samples, one being from 45 teachers and the other from 32 foreign teachers both with their 3rd year students. It then compares this with the outcome of the student academically to find the correlation. The study suggests that behaviour does have an affect on the student's performance.

2. The Effect of Teacher Behaviour on Students Motivation and Learning Outcome

This paper talks about the effect of teacher behaviour and the student's motivation to learn. The data is collected from a research regarding the same idea. It

also investigates the teaching styles of teachers like how supportive or controlling a teacher is.

3. The Effect of Teacher Interpersonal Behaviour on Students' Subject Motivation

This study investigates the effect of the teacher's behaviour on their student's motivation for the subject they are teaching. The behaviour of the teacher was measured using 2 independent variables; Influence and Proximity. The data is taken from 52 third-year classes taught by 32 individual teachers. The study finds that interpersonal behaviour does have a significant effect on the motivation of students.

4. Impact of Course Length on Student Learning

This study goes into the impact of the length of a course to determine the correlation with student learning. It uses data from a database with about 45000 observations. The study discovers that longer courses results in higher grades. This benefit diminishes once the extension is longer than 4 weeks.

5. Increased Course Structure Improves Performance in Introductory Biology

This study tries to investigate whether a highly structured course would improve student performance in introductory biology courses. This course is structured in a way where it implements quizzes, active learning activities and practice exams weekly. The performance of this is then compared with a less structured course based more on lecturing. From the study's findings, the active learning activities had a large impact on the performance.

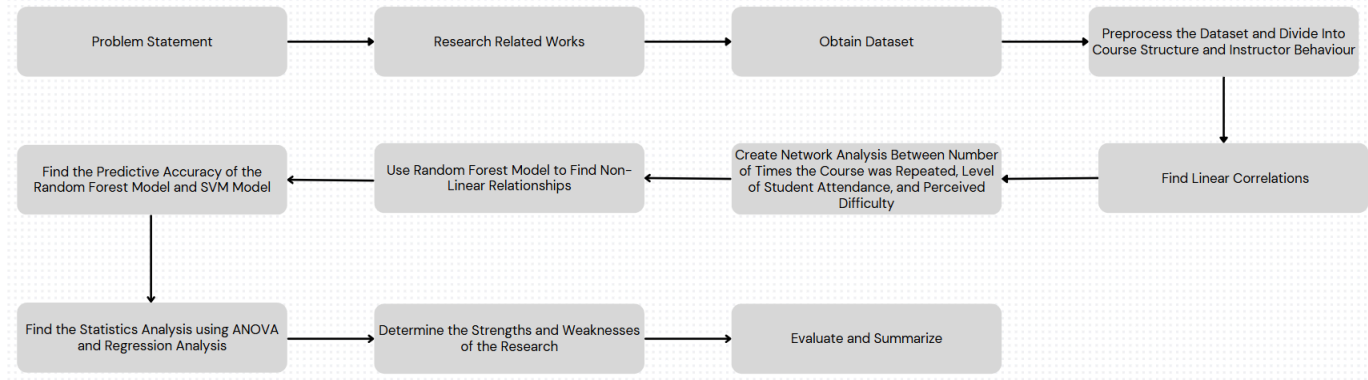
2.3 Research Gap

The research gap between the studies investigated in the related studies section and this report is that those studies specifically investigate one element. The studies previously investigated either talk about either course structure or teacher behaviour, not both. This may have a significant impact on the finding, as it is easier to find which factor has a larger impact on student performance.

Chapter 3. Methodology

3.1 Flowchart

Chart 1. Flowchart of the Research



3.2 Dataset

The dataset is obtained from students in Gazi University (Turkey) and has 5820 evaluation scores. This dataset consists of 33 columns where 12 columns focus on the course structure, 16 columns on instructor behaviour, and there are 3 columns that act as the output or outcome which are “repeat”, “attendance”, and “difficulty”.

Table 1. Dataset Variables Information

Field Name	Description	Values
instr	The instructor where the values are	{1,2,3}
class	The course identifier where the values are {1-13}	{1-13}
repeat	The number of times that students repeated that specific course	{0,1,2,3,...}
attendance	The level of attendance that the student have throughout the semester	{0,1,2,3,4}
difficulty	The level of difficulty that students perceive of the course	{1,2,3,4,5}
Q1	How complete the content, methodology, and evaluation system of the course was provided during the beginning of the semester	{1,2,3,4,5}
Q2	Whether the objective of the course was explained during the start of the semester	{1,2,3,4,5}

Q3	Whether the credits received was worth it from the course	{1,2,3,4,5}
Q4	Whether the teachings of the course was aligned based on the syllabus from the start of the semester	{1,2,3,4,5}
Q5	How satisfactory was the course activities such as discussions, assignments, and studies	{1,2,3,4,5}
Q6	School resources such as textbook is up to date	{1,2,3,4,5}
Q7	Give opportunities to do field work, applications, laboratory, discussion, and other ways of studies.	{1,2,3,4,5}
Q8	Quizzes, assignments, exams, and projects play a major role in helping to learn	{1,2,3,4,5}
Q9	I am eager and enjoy being active and participating on the lectures	{1,2,3,4,5}
Q10	The course met my initial expectation of the course at the end of the year	{1,2,3,4,5}
Q11	The course was highly relevant and contributes a lot to my professional growth	{1,2,3,4,5}
Q12	The course opened up a new perspective on the world and life	{1,2,3,4,5}
Q13	The instructor shows that their knowledge of the course is relevant and up-to-date	{1,2,3,4,5}
Q14	The instructor was prepared for all class sessions	{1,2,3,4,5}
Q15	The instructor follows the lesson plan throughout the course	{1,2,3,4,5}
Q16	The instructor is dedicated and communicates well throughout the course	{1,2,3,4,5}
Q17	The instructor always arrives on time to class	{1,2,3,4,5}
Q18	The instructor delivers the course in a clear way and is easy to follow.	{1,2,3,4,5}
Q19	The instructor makes full use of the class hours efficiently and effectively	{1,2,3,4,5}
Q20	The instructor is eager to teach and help the students	{1,2,3,4,5}
Q21	The instructor is approachable to the students	{1,2,3,4,5}
Q22	The instructor is open-minded and respectful of the	{1,2,3,4,5}

	student's views about the course	
Q23	The instructor consistently encouraged students to participate and engage in the course	{1,2,3,4,5}
Q24	The instructor assigns homework and projects that are relevant and meaningful to the course	{1,2,3,4,5}
Q25	The instructor responds to questions about the course both inside and outside of class sessions	{1,2,3,4,5}
Q26	The way the instructor evaluates the assignment is aligned with the course	{1,2,3,4,5}
Q27	The instructor discusses the exam thoroughly with the students	{1,2,3,4,5}
Q28	The instructor treats all students fairly and in an objective manner	{1,2,3,4,5}

3.3 Preprocessing

The code was preprocessed by checking for any null values, where after checking shows that the dataset has no null value present indicating that the next step can be taken. The dataset itself has 5820 records showing that it is less likely to overfit and show any signs of data imbalance. The field names were divided into course structure which is from Q1 to Q12 and instructor behaviour which is from Q13 to Q28, this was done so that the effects from each of course structure and instructor behaviour can be observed clearly. During the processing of the dataset from the models explained below, no errors of data imbalance occurred indicating that the process was a success.

3.4 Models and Techniques

3.4.1 Linear Correlation

The correlation analysis is used to find linear relationships between Q1 to Q28 and repeat, attendance, and difficulty. By being able to find these linear relationships, it can help in interpreting the data and find strongly correlated fields (Hassan, 2024).

3.4.2 Network Analysis

Network analysis examines the relationship between the variables by visualizing the connections between the multiple variables. In this script, it analyses the relations between nb.repeat, attendance, and difficulty of the course, by understanding their connection and checking how the variables influence each other. By visualizing the connection as a network

the correlation value will show how much a variable influences another. Not only that, degree centrality shows how much a variable influences the network, to see which variable has the central role (Smiraglia, 2015)

3.4.3 Random Forest

Random forest analysis creates multiple decision trees during training, where each of those trees has random inputs of data to create variability. Having a number of these trees reduces the likelihood of overfitting and gives an edge of increasing the predictive power. The random forest analysis helps in finding non-linear relationships thus helping in interpreting the data where linear correlation analysis may not find (GeeksforGeeks, 2024).

3.4.4 Linear Regression

Linear regression is used to predict the nb.repeat which is continuous, to see how often students repeat the course. Linear regression is being used in our model to focus on the course-related features in Q1-Q12 to see which aspect of the course plays a significant role in the repetition of students. The model is then evaluated with mean squared error or MSE so that the prediction can be compared with the actual values so that it can be seen whether the model is accurate or not (GeeksforGeeks, 2024).

3.4.5 Support Vector Machines

Support Vector Machine (SVM) is used for variables that are categorical such as attendance and difficulty since it is easier to classify them. It is used because it can separate different groups such as people with high attendance and people with low attendance in complex data with many variables. It uses both course-related and instructor-related questions to see the wider image of what affects the students rather than only one variable-related question, instructor, or course (GeeksforGeeks, 2024).

3.4.6 Performance Metrics

Performance metrics that are used in this code are accuracy which are used to compare how the models work. Accuracy in this case will help with how well the model of Random Forest and SVM predicts the output. On the other hand, MSE will measure the error of nb.repeat which is a continuous outcome. It decides which model gives the best result (GeeksforGeeks, 2024).

Chapter 4. Discussion

4.1 Result and Data Interpretations

Figure 1. Linear Correlations between course structure and number of times the course was repeated

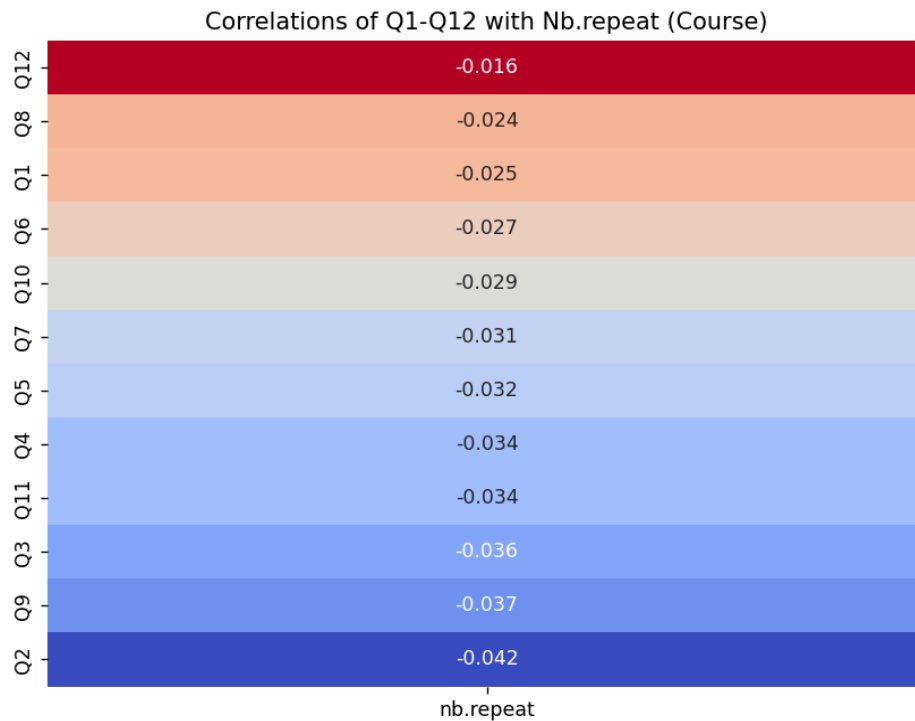


Figure 2. Linear Correlations between instructor behaviour and number of times the course was repeated

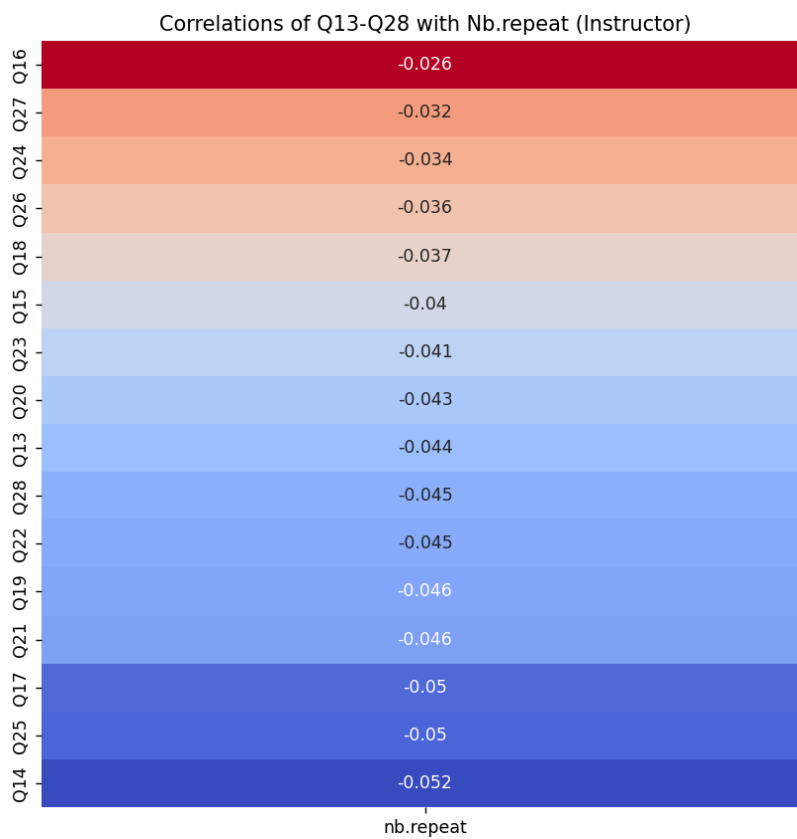


Figure 3. Linear Correlations between course structure and the student attendance

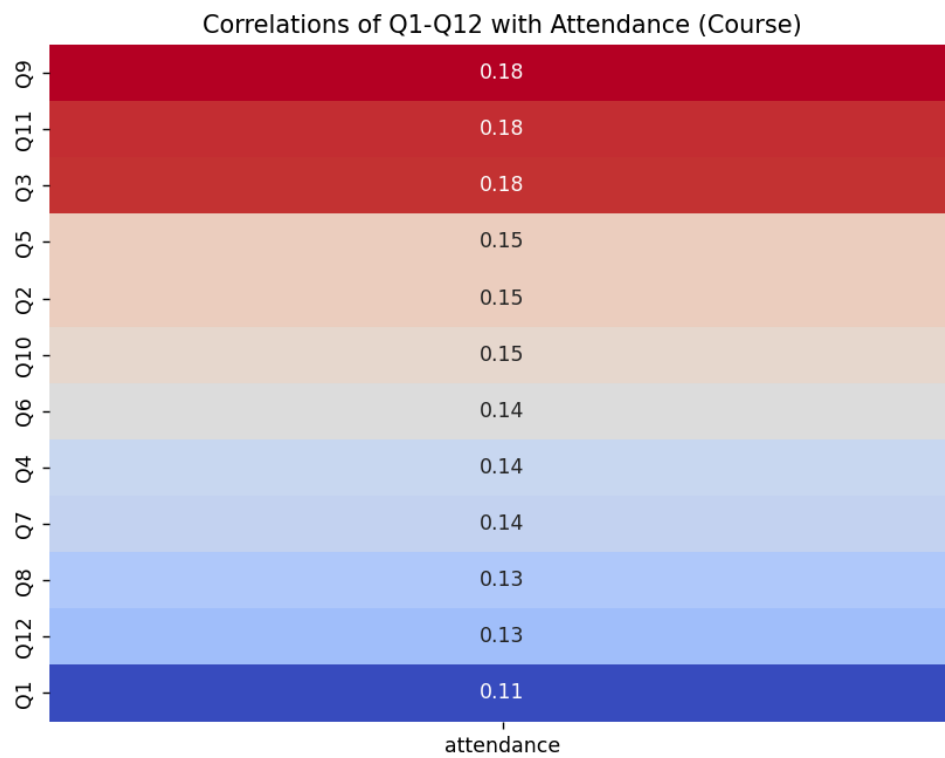


Figure 4. Linear Correlations between instructor behaviour and student attendance

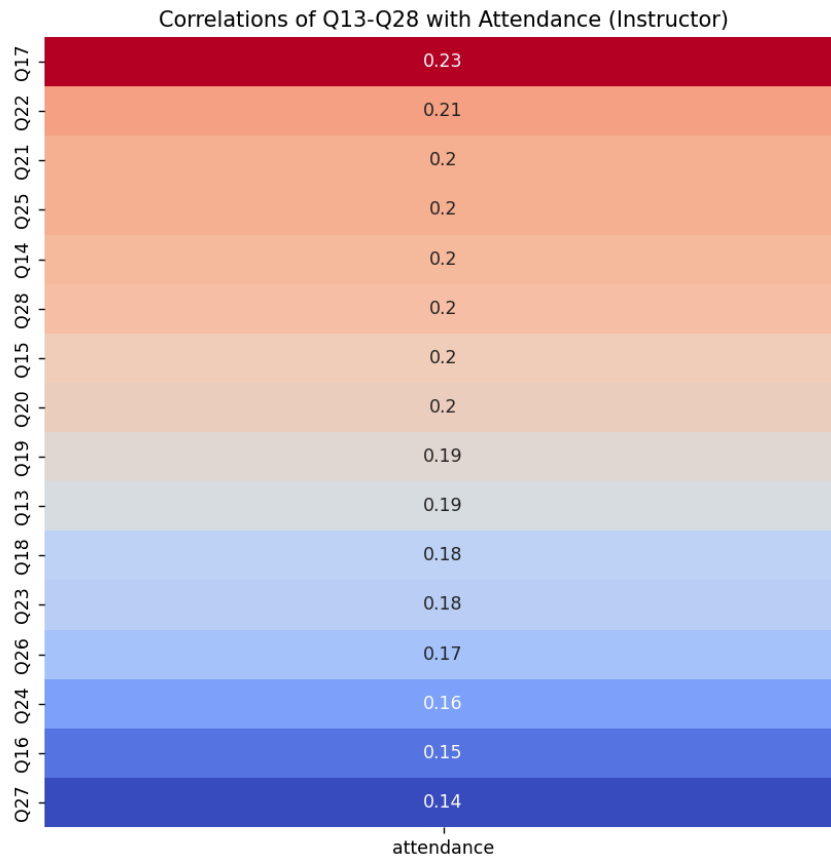


Figure 5. Linear Correlations between course structure and the course difficulty

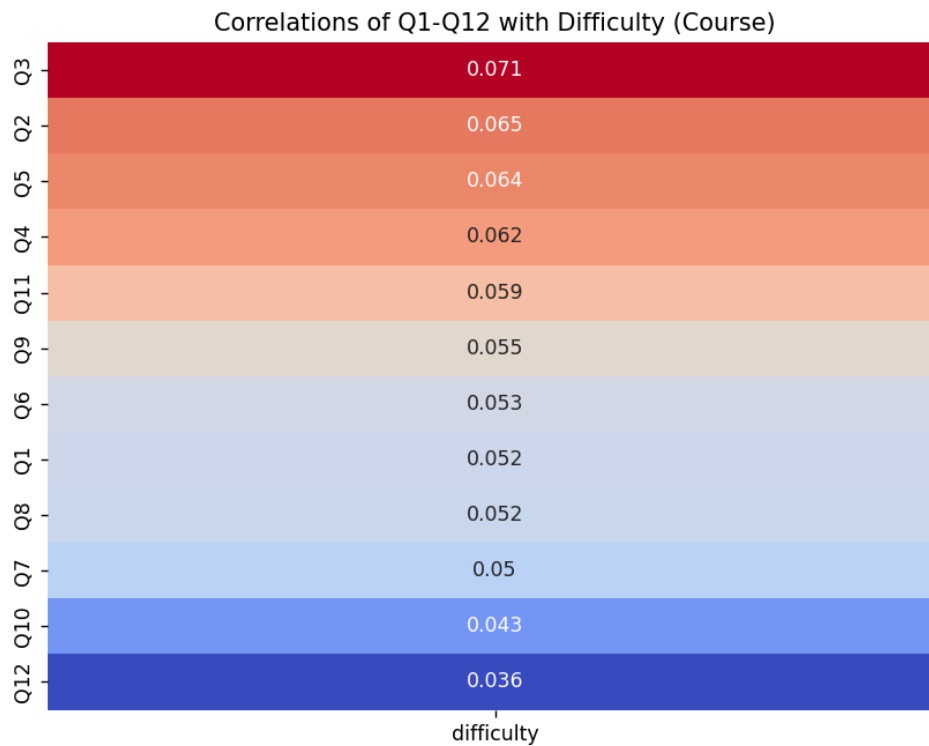
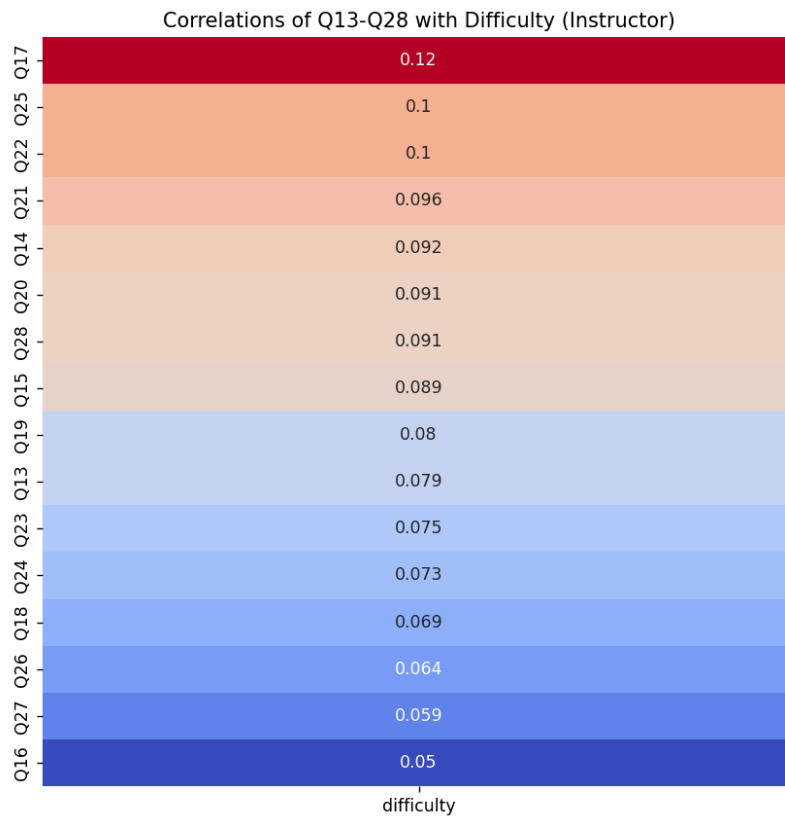


Figure 6. Linear Correlations between instructor behaviour and the course difficulty



4.1.1 Linear Correlations

Figure 1 up to Figure 6 shows the linear relationships between course structure and instructor behaviour towards the number of times the course was repeated, the student's attendance level, and how difficult the students perceived the course. The measurement numbers shown above use correlation coefficient to calculate the correlation linearly.

Figure 1 and 2 shows the relationships between course structure and instructor behaviour towards the number of times the course was repeated. Both variables show a negative correlation coefficient, indicating that as the values of the course structure and instructor behaviour increases, the repetitions of courses decreases. However, the two figures also show a very weak correlation, where the highest correlation for course structure is Q2 (Whether the objective of the course was explained during the start of the semester) at -0.042 and for instructor behaviour is Q14 (The instructor was prepared for all class sessions) at -0.052. With the very weak correlation between these two variables, indicates that whether these values decrease or increase, gives little to no linear impact towards the number of times the course was repeated.

Figure 3 and 4 shows the relationships between course structure and instructor behaviour on the level of the student's attendance. The two variables show a positive

correlation, which indicates that as the values of the two variables increase, the likelihood of students to attend the course increases as well. The highest correlation coefficient from the course structure is Q9 (I am eager and enjoy being active and participating in the lecture) at 0.18 and for instructor behaviour is Q17 (The instructor always arrives on time to class) at a higher number of 0.23. Both of these variables explain that when the class is more engaging and the instructor is punctual has the highest effect on the level of attendance that the students has, however the numbers show that it has only a slight effect due to the low value of the correlation coefficient at only 0.18 and 0.23. It is reasonable that none of the values show a high correlation coefficient as the engagement of the class or instructor's punctuality on the student's attendance does not have a high effect on every student evenly. This slight effect does not indicate that the values are pointless, rather it shows that to increase the student's attendance, there needs to be a combination of multiple high value scores for the course structure and instructor behaviour.

Figure 5 and 6 shows the relationships between course structure and instructor behaviour towards how difficult the students perceived the course. Both variables show a positive correlation indicating that as the correlation coefficient increases, then the perceived difficulty of the course increases as well. The highest value from course structure is Q3 (Whether the credits received was worth it from the course) at 0.071 and for instructor behaviour is Q17 (The instructor always arrives on time to class) at 0.12. Although both correlation coefficient values are small, it still signifies a positive correlation that when the value of the course structure and instructor behaviour increases, the difficulty of the course is higher as well. This goes against the hypothesis as the increased value of both variables was hypothesised to decrease the difficulty and help learning effectiveness of the students, however the result shows that it says otherwise. The cause may have been that as both variables value increases meaning that example, the worth of the credits, clear course objectives, discussions/assignments, punctuality of the instructor, etc may have given more pressure towards the students which contributed to the perceived difficulty of the course, although it is slightly.

Overall, course structure and instructor behaviour shows only a slight linear relationship with the number of times the course was repeated, the level of attendance that the students have, and the perceived difficulty of the courses. The result aligns with the hypothesis in two aspects:

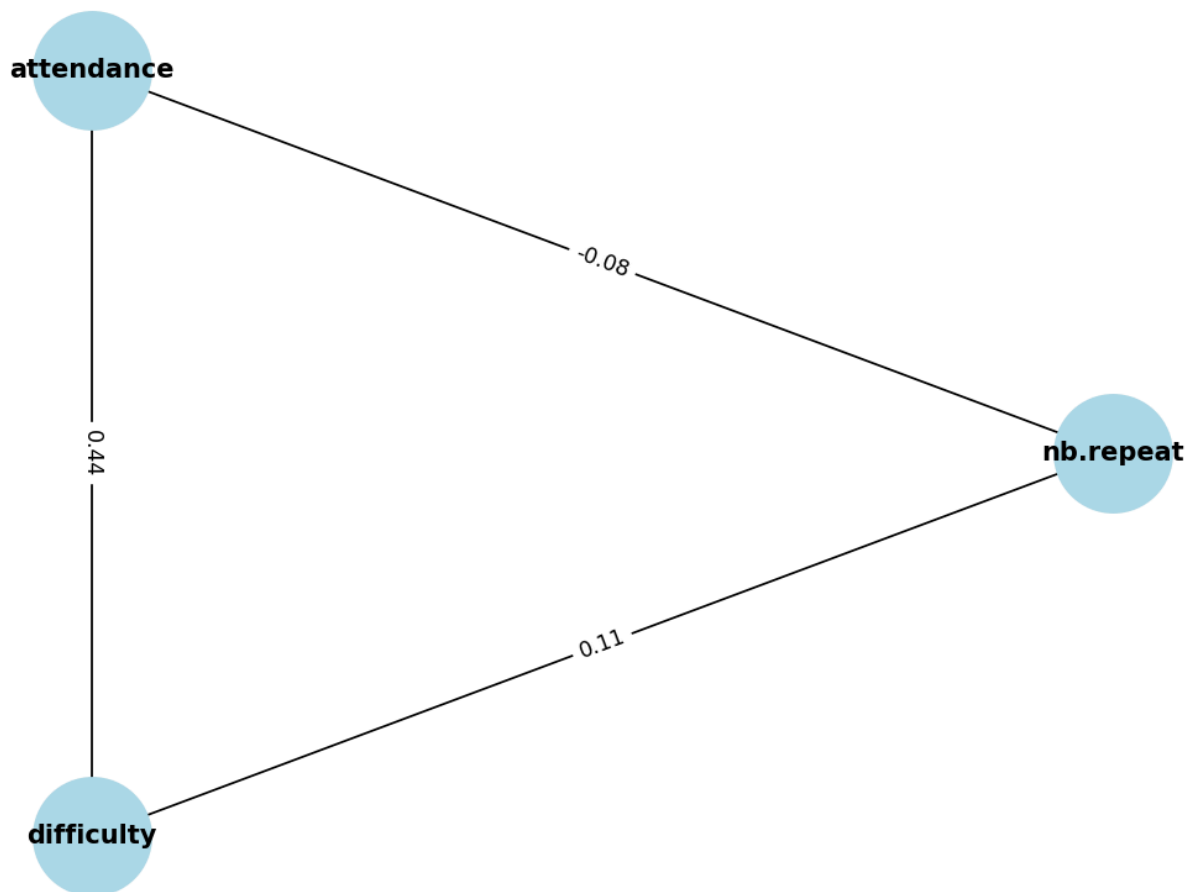
- The correlation coefficient of the course structure and instructor behaviour towards the number of times the courses was repeated shows that as the value decreases, the

frequency of course repetition increases. Although the value is very small however it still signifies a negative correlation showing that it aligns with the hypothesis, although very slightly.

- The correlation coefficient of the course structure and instructor behaviour towards attendance shows that as the value increases, the level of attendance that students have also increases. The value shows a slight to moderate effect indicating that it still aligns with the hypothesis.

However the hypothesis was wrong in one aspect, where the increased value of the course structure and instructor behaviour did not decrease the difficulty of the course, rather it increased it. Although the value is small, it still shows that it goes against the hypothesis.

Figure 7. Network Analysis Between Attendance, Difficulty, and Repeated Courses



4.1.2 Network Analysis

The diagram above shows the result of the network analysis between attendance, course difficulty, and repetition of the course. The three values show the relationships between each dependent variable. The three values are:

- Attendance - Repetition: -0.08

This means that there is a small negative correlation between attendance and the frequency of repetition. The value is very small however it still signals that when the attendance is lower then there is a slightly higher chance that the number of times the course was repeated would be higher.

- Repetition - Difficulty: 0.11

This means that there is a slight positive correlation between the frequency of repetition and the difficulty of the course. This shows that when the perceived course difficulty increases, it contributes to the frequency of the repetition of the course.

- Difficulty - Attendance: 0.44

This means that there is a moderate positive correlation between the difficulty of the course and the level of attendance that the student has. The value shows that when the attendance increases, it is likely that the perceived difficulty of the course increases as well.

The network analysis has similarities with the previous linear correlation analysis. It can be observed from the first value, the relationship between attendance and the frequency of repetition shows a very low negative correlation. This case also happens in the linear correlation analysis where there is also a low negative correlation between the course structure and instructor behaviour with the number of times the course was repeated. This strengthens the findings that there is very little impact towards the frequency of repetition.

The second value of 0.11 shows the slight positive correlation between the perceived difficulty of the course and the frequency of repetition. This also supports the findings from the linear correlation analysis and the first point (attendance and the frequency of repetition), showing a small correlation indicating that there is only a slight impact when it comes to affecting the frequency of repetition of the course.

The third value of 0.44 shows the relationship between the perceived difficulty of the course and the level of attendance that the student has; it appears to show a moderate positive correlation. This result is similar to the previous linear correlation analysis where it shows how course structure and instructor behaviour has a positive correlation with the perceived difficulty. This goes against the hypothesis where it is rather expected that when a student has a higher level of attendance, the perceived level of difficulty would be lower. However, the results show that when a student has a higher level of attendance, it would also increase the perceived level of difficulty.

Figure 8. Course Feature Importance Towards Repeated Courses

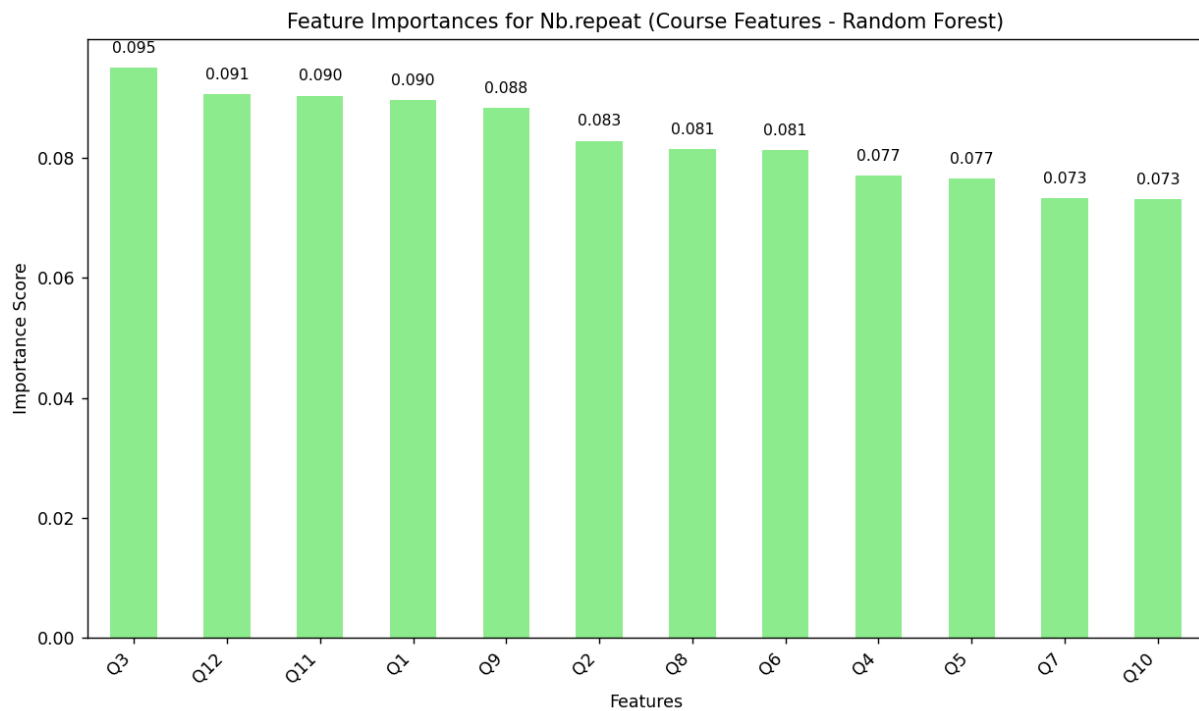


Figure 9. Instructor Behaviour Importance Towards Repeated Course

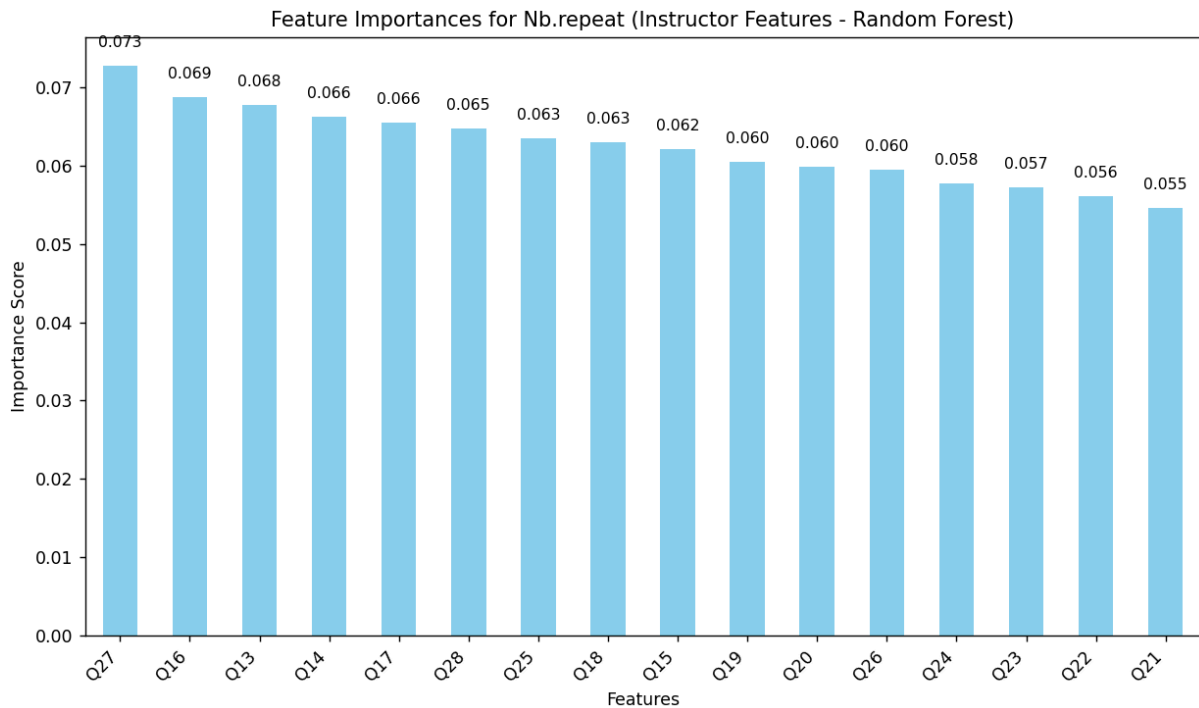


Figure 10. Course Feature Importance Towards Attendance

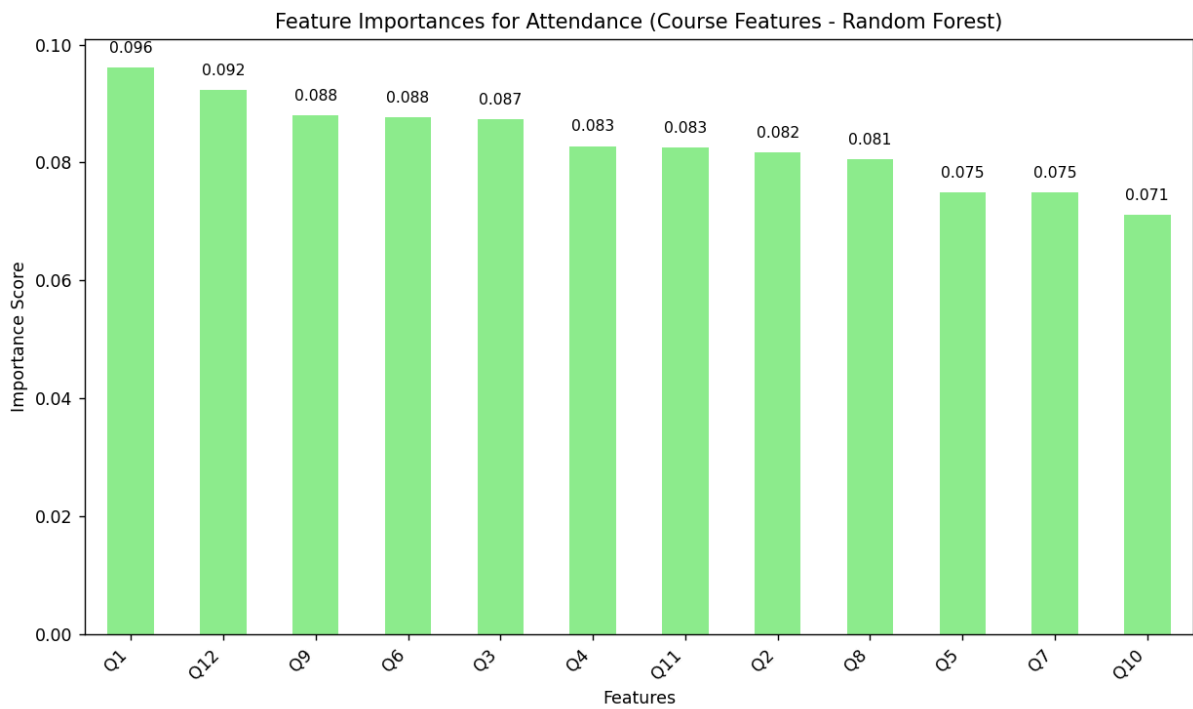


Figure 11. Instructor Behaviour Importance Towards Attendance

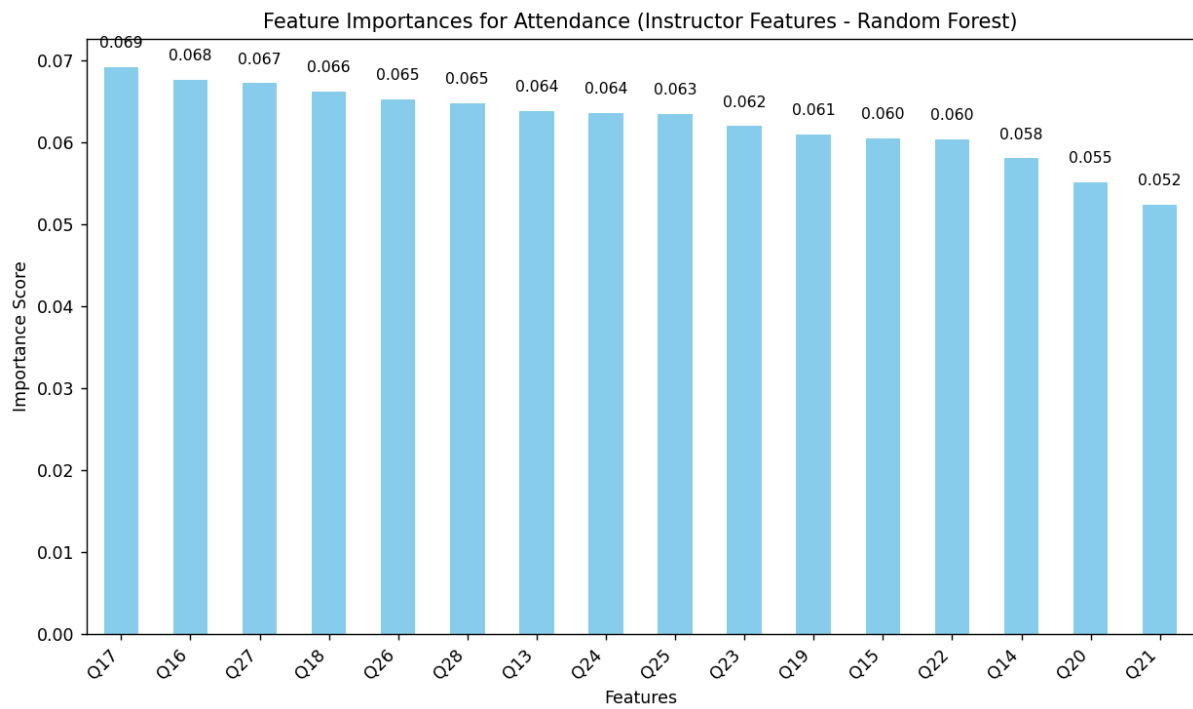


Figure 12. Course Features Importance Towards Difficulty

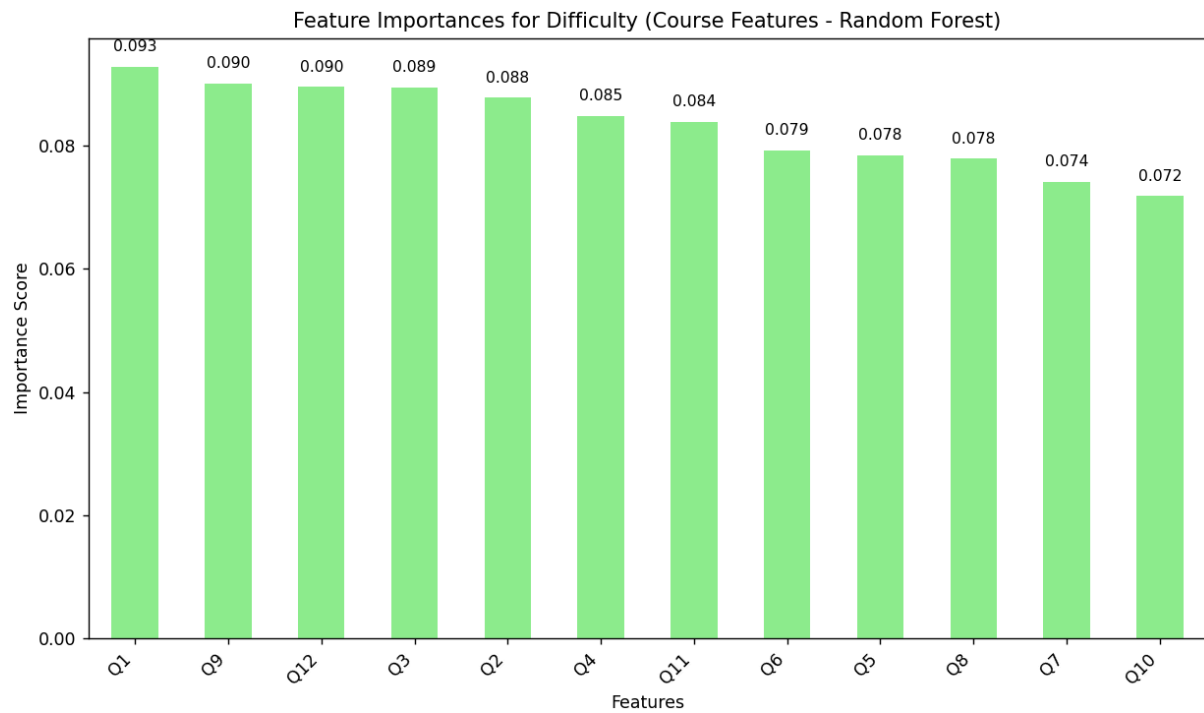
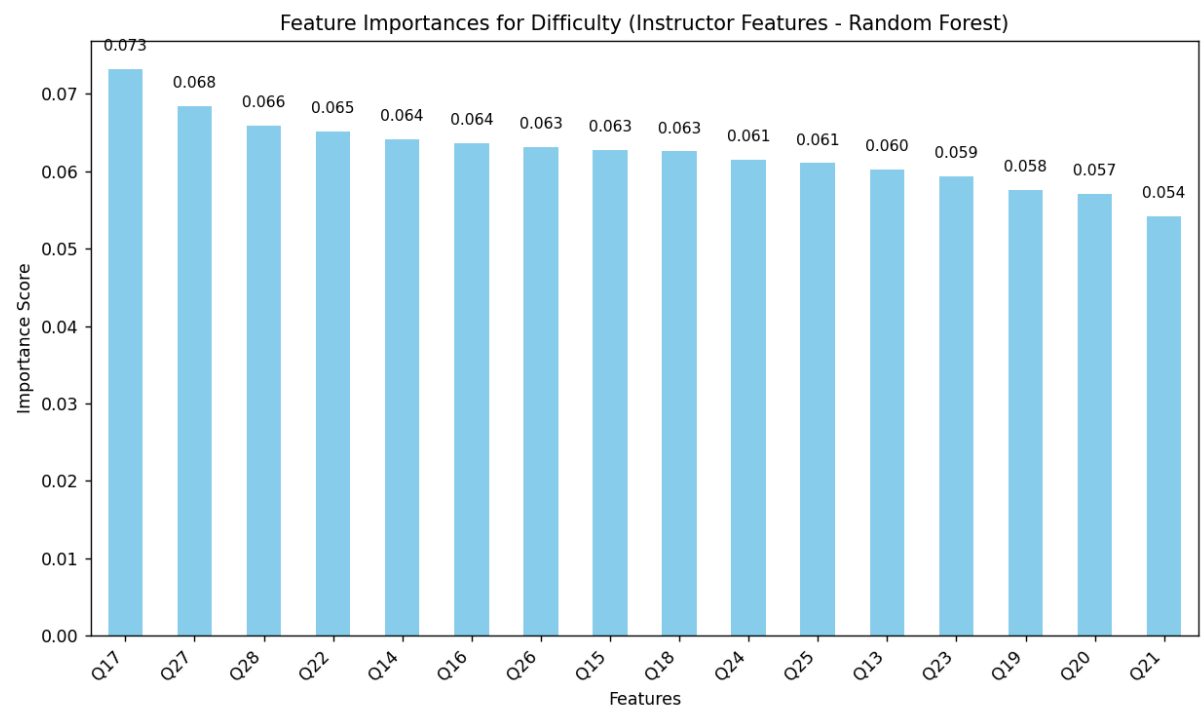


Figure 13. Instructor Behaviour Importance Towards Difficulty



4.1.3 Random Forest Analysis

Figure 8 to Figure 13 shows the non-linear relationships using random forest between course structure and instructor behaviour towards the number of times the course was repeated, the student's attendance level, and how difficult the students perceived the course.

Figure 8 and 9 shows the correlation between course structure and instructor behaviour towards frequency of course repetition. The highest importance score for course structure is Q3 (Whether the credits received was worth it from the course) at 0.095 and for instructor behaviour is Q27 (The instructor discusses the exam thoroughly with the students) at. The two relationships show a weak positive correlation with the repetition, although it is weak, it still signifies that when the credits are worth and the instructor gives a clear explanation of the exam, it would give a slight impact on the frequency of repetition.

The results in this non-linear relationship analysis shows differences compared to the linear relationships. This difference indicates that there are some correlations that the linear analysis did not pick up. In the linear correlation analysis between course structure and frequency of repetition, the highest value is Q2 (Whether the objective of the course was explained during the start of the semester) and Q3 is ranked 3rd, however in figure 8 shows that Q2 is ranked 6th and Q3 as 1st. The linear correlation between instructor behaviour and frequency of repetition, the highest value is Q14 (The instructor was prepared for all class sessions) and Q27 is ranked 14th, however in figure 9 shows that Q14 is ranked 4th.

Figure 10 and 11 shows the correlation between course structure and instructor behaviour towards the attendance of students. The highest importance for course structure is Q1 (How complete the content, methodology, and evaluation system of the course was provided during the beginning of the semester) at 0.096 and for instructor behaviour is Q17 (The instructor always arrives on time to class) at 0.069. Both of these values show a weak positive correlation indicating that the importance is not that strong although it is calculated as the most important from the random forest program. It shows that the clear content explanation in the beginning of the semester and the attendance of the instructor has the highest effect in terms of student's attendance.

From the linear correlations analysis, the highest value that shows the strongest correlation for course structure is Q9 (I am eager and enjoy being active and participating in the lectures) and it is placed at 3rd for the highest importance, while Q1 is ranked last in figure 3. The highest value for instructor behaviour is Q17 which is the same for the random forest relationship, this shows that both correlations agree that Q17 is the strongest indicator for a high student attendance level.

Figure 12 and 13 shows the correlation between course structure and instructor behaviour towards the perceived difficulty of the course. The course structure feature with the highest importance is Q1 (How complete the content, methodology, and evaluation system of the course was provided during the beginning of the semester) at 0.093 and instructor

behaviour is Q17 (The instructor always arrives on time to class) at 0.073. Both shows a positive correlation with the perceived difficulty of the course.

It is noticeable how figure 10 and 11 with figure 12 and 13 shows the same highest feature importance. From the linear correlation, instructor behaviour for perceived difficulty and attendance shows Q17 having the highest correlation coefficient, which is similar with the random forest classifier showing that Q17 also has the highest value in terms of instructor behaviour. However for course structure, there are no visible significant similarities between course structure in the linear correlation and random forest towards perceived difficulty. In the linear correlation, Q3 (Whether the credits received was worth it from the course) has the highest correlation coefficient while Q1 is at 8th place, and in the random forest, Q3 is 4th place. It is noticeable how figure 10 and 11 with figure 12 and 13 shows the same highest feature importance. This is similar to the previous network analysis, where it is shown that difficulty and attendance has the highest correlation at 0.44, a moderate value however showing a more significant sign that there is a relationship between the two.

The correlations above show factors which are regarded as the highest importance, however the value of the other features of every factor are not far off, indicating that even though the highest valued features have the highest importance, it is not that significant. Although it is not that significant, it does not necessarily mean that the importance scores are pointless, but shows that it would need multiple independent variables to actually have an impact.

Table 2. Predictive Accuracy for Number of Times the Course was Repeated

<i>Model</i>	<i>Accuracy</i>
<i>Random Forest Accuracy for repetition (Course Features)</i>	0.84
<i>Random Forest Accuracy for repetition (Instructor Features)</i>	0.84
<i>SVM Accuracy for Nb.repeat (All Features)</i>	0.84

4.1.4 Performance Evaluation

Table 2 shows a high predictive accuracy score for the random forest and SVM model, indicating that there are strong correlations which helps the model to predict the right frequency of course repetition. Analyzing the accuracy of each model will help to understand the strength of the correlations, outside of only looking at linear correlations. Both random forest models for course structure and instructor behaviour have the same accuracy score

indicating that both have an equal contribution in the predictive accuracy for the model. This equal contribution can also be backed by the SVM accuracy which combines both features and still has the same predictive accuracy, showing that there is minimal interaction between the two features. The linear regression MSE (Mean Squared Error) measures the average squared difference between the actual and predicted values (GeeksforGeeks, 2024).

The data above only shows the predictive accuracy for frequency of repetition because the random forest and SVM models had difficulty predicting the attendance and perceived difficulty. Both having a predictive accuracy below 50% indicates that the model is inaccurate thus a conclusion cannot be drawn.

4.2 Statistical Analysis

Regression Analysis and ANOVA Analysis

Regression analysis is a method used to estimate the relationship between dependent variables and one or more independent variables (GeeksforGeeks, 2024). In this study, regression analysis helps estimate the relationship of how instructor behaviour and course structure influence learning effectiveness which are: Attendance, course, repetition, and the difficulty for students. This analysis will use both course-related features and instructor-related features. Visuals such as the scatter plot diagram help give insights into which factors have a meaningful impact on learning effectiveness and also give insight into where factors have less meaningful impacts.

Furthermore, ANOVA (Analysis of Variance) is used to see the mean values of the course structure and instructor behaviour to see if there are meaningful differences between the groups. To see whether or not there is a change in instructor behaviour and course structure will it be different for groups of students with different answers?

Attendance

```
Linear Regression MSE for Attendance (Course Features): 2.11
Linear Regression R^2 for Attendance (Course Features): 0.0375
Linear Regression MSE for Attendance (Instructor Features): 2.07
Linear Regression R^2 for Attendance (Instructor Features): 0.0567
```

```
Course Structure: F-statistic = 578.4735, p-value = 0.0000e+00
Instructor Features: F-statistic = 1098.4687, p-value = 0.0000e+00
```

Using Anova Analysis to course structure reveals F-statistic = 578.47, p-value \approx 0.0000 and instructor-related features which show F-statistic = 1098.47 and p-value \approx 0.0000. This shows that both instructor-related features and course-related features have a major relationship with attendance. However, regression shows restriction in the success of predicting attendance as a factor indicator of students' learning effectiveness. It is shown that course-related features gave R^2 value of 0.037 and instructor-related features gave R^2 a value of 0.057 which is a slight increase in value in comparison with course-related features. This shows that there is a very small effect that the course-related and instructor-related plays in the factor of attendance. However, there are specific factors that stood out from the rest which are Q9 and Q17. Q9 (Students enjoyment and engagement in lectures) and Q17(Instructor punctuality). Even with those factors showing a slight positive impact, their influence is too small to be significant in predicting the attendance of students. We believe that there are outside factors that might have a bigger impact on the attendance. The analysis shows low support of our hypothesis that course structure and instructor behaviour positively impact the attendance of students. Even though there are factors such as Q9 and Q17 that show a small positive effect, the overall impact is too small to show that it positively affects student attendance.

Course Repetition

```
Linear Regression MSE for Nb.repeat (Course Features): 0.31
Linear Regression R^2 for Nb.repeat (Course Features): -0.0004
Linear Regression MSE for Nb.repeat (Instructor Features): 0.31
Linear Regression R^2 for Nb.repeat (Instructor Features): -0.0011
```

Turns out regression analysis shows that there is a very weak correlation in both course-related features which have a value of -0.0004 and instructor-related features which have a value of -0.0011. Since $R^2 = 0$ means that the model is not doing any better than the baseline prediction, this shows that the model shows a very weak relation between the two factors and course repetition. Though there are some of the weaker relations that reduced the repetition rate such as Q12 (The course broadened perspective) and Q16 (The instructor communicated effectively), this is still too small to create an impact in the effect of course repetition. Again, this shows that there are external factors that affect course repetition rather than course features and instructor features. This goes against our hypothesis of course structure and instructor behaviour reducing the course repetition since the result doesn't

support our hypothesis. It shows that these factors play a very small role in whether or not a student repeats the course or not.

Perceived Difficulty

```
Linear Regression MSE for Difficulty (Course Features): 1.83  
Linear Regression R^2 for Difficulty (Course Features): 0.0095  
Linear Regression MSE for Difficulty (Instructor Features): 1.80  
Linear Regression R^2 for Difficulty (Instructor Features): 0.0246
```

Regression analysis also showed a weak relationship between course-related features and instructor-related features. Course-related features have R^2 a value of 0.0095, and instructor-related features have a R^2 value of 0.0246. This shows that there is a very low relation between the factors and how it affects the difficulty of learning for students. However, there were also factors that made students perceive that there is a higher difficulty in the course based on Q3 (Whether the credits received were worth it from the course) and Q17 (The instructor always arrives on time to class). This analysis contradicts the hypothesis that the course structure and instructor behaviour will decrease the perceived difficulty. This analysis suggests that the improvement of these factors will increase the difficulty for students.

Conclusion

Regression and ANOVA analysis shows that instructor behaviour and course structure have a small impact on learning effectiveness. Though attendance is slightly affected by the two factors, the other two which are course repetition and perceived difficulty are mostly not affected. These findings mostly contradict our hypotheses, which show that the course structure and instructor behaviour play a smaller role in these areas than expected. These results also show the complexity of learning effectiveness, such as personal behaviour, factors, social surroundings, and available supply that is not part of this dataset.

4.3 Strengths

Comprehensive analysis

This analysis covers many aspects of learning effectiveness such as attendance, course repetition and difficulty and includes using a dataset with course-related features (Q1-Q12) and instructor-related features (Q13-Q28) which covers the factors influencing learning effectiveness. This many datasets allow for a precise methodical examination of the factors that influence learning effectiveness.

Application of Statistical Metrics

The purpose of R^2 is to show the relationship between the models and how good the model is for the data. Mean Squared Error (MSE) is used to see the difference between the predicted results are first the actual results, and second to see how precise the models are.

Graphical Visualization

The code produces visual outputs such as graphs and charts to point out the relationship between the variables. These visuals help to present which variable is most influential towards learning effectiveness, such as which specific instructor's behaviour affects student attendance the most.

More Than One Analytical Approach

Linear regression is used to find and interpret the linear relationships between the variables in the dataset. In contrast, random forest finds non-linear relationships and patterns that are harder to detect by linear regression.

4.4 Weaknesses

Potential Biased Responses

The dataset might have biased responses from the participants, which may reduce the accuracy and reliability of the study. The reason for bias may come from student's different experiences with the instructor and different courses. Due to this the answer that was given by the students may be influenced by their own feelings or opinion. The data might not present the actual relationship between the different factors.

Limited Scope of External Factors

The dataset doesn't cover external factors such as students' economical situation, family situation, mental well-being, or their socioeconomic background. These factors play a major role in attendance, repetition, and academic performance in general. Since this dataset doesn't have such factors it makes the depth of this analysis less complete.

Narrow Data Collection Source

The dataset is taken only from one university that is Gazi University from Ankara, this may not present the broader educational situation. There are differences in course structure and instructor's behaviour in different universities therefore this analysis may not be able to represent a wider scale.

Code Limitation

The code doesn't handle outliers or missing values, which can affect the reliability of the result. Outliers that are not handled properly may mess up the result, which leads to inaccurate analysis and conclusion. Missing values can also decrease the reliability in the analysis.

Chapter 5. Conclusion

The strongest variable for the correlation between course, repeat, attendance and difficulty are Q3. Q3 is a variable that measures if the credits received were worth it from the course. Q3 seems to be the strongest in this case because of its positioning in Figures 1, 3, and 5. Q3 is always in the top 3 highest correlations in the three figures mentioned. It is the highest correlation outright in figure 5 and is tied for first with 2 other variables. The strongest variable for correlation between instructor, repeat, attendance and difficulty are Q17. It is a variable that measures if the instructor arrives on time. The reason that Q17 is the strongest variable in this case is similar to Q3, it is almost always the highest value in Figures 2, 4, and 6. Specifically, Q17 has a value of 0.05 compared to the highest 0.052 in Figure 2. While in Figure 4, it has outright the highest value of 0.23. Finally, in Figure 6, it has a value of 0.12 making it almost always the highest value throughout the 3 Figures.

Although the results explained before may seem promising, there are some issues. Specifically, results from Regression and ANOVA analysis show that even features like Q3

and Q17 do not have a large impact. This means that the results or proof of course structure and instructor behaviour do not play as large of a role as the hypothesis shows. This shows the complexities of analysing the effect of instructor behaviour and course structure on performance, too many outside factors could be at play affecting the results further.

Overall, even though the most important variables that affect learning performance were identified. They have quite a small impact in the grand scale, this means that on their own these variables need to work together to have a cumulative effect on learning performance. Another thing to note regarding weaknesses is the dataset, the dataset may be biased as not all students might respond fairly or unbiasedly. It also may not be accurate due to other outside factors that are not instructor behaviour or course structure.

Reference

- den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal Teacher Behaviour and Student Outcomes. *School Effectiveness and School Improvement*, 15(3–4), 407–442.
<https://doi.org/10.1080/09243450512331383262>
- Hein, V. (2012). The effect of teacher behaviour on students motivation and learning outcomes: a review. *Acta Kinesiologiae Universitatis Tartuensis*, 18, 9.
<https://doi.org/10.12697/akut.2012.18.02>
- The Effect of teacher interpersonal behaviour on Students' Subject-Specific Motivation on JSTOR. (n.d.). www.jstor.org.
<https://www.jstor.org/stable/23870661>
- Austin, A., & Gustafson, L. (2020). Impact of Course Length on Student Learning. *JOURNAL OF ECONOMICS AND FINANCE EDUCATION*, 5(1),
https://www.montgomerycollege.edu/_documents/offices/academics-affairs/covid-19-contingency-plans/resources/2020-08-20-impact-of-course-length-on-student-learning.pdf.
- Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education*, 10(2), 175–186. <https://doi.org/10.1187/cbe.10-08-0105>
- Hassan, M. (2024, November 18). *Correlation Analysis - types, methods and examples*. Research Method. <https://researchmethod.net/correlation-analysis/>
- Smiraglia, R. P. (2015). Empirical techniques for visualizing domains. In *Elsevier eBooks* (pp. 51–89). <https://doi.org/10.1016/b978-0-08-100150-9.00004-3>
- GeeksforGeeks. (2024, December 11). *Random Forest algorithm in machine learning*. GeeksforGeeks.
<https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>
- GeeksforGeeks. (2024, October 23). *Linear Regression in Machine learning*. GeeksforGeeks. <https://www.geeksforgeeks.org/ml-linear-regression/>

GeeksforGeeks. (2024, October 10). *Support Vector Machine (SVM) algorithm*.

GeeksforGeeks.

<https://www.geeksforgeeks.org/support-vector-machine-algorithm/>

GeeksforGeeks. (2024, August 13). *Mean squared error*. GeeksforGeeks.

<https://www.geeksforgeeks.org/mean-squared-error/>

GeeksforGeeks. (2024, August 13). *What is Regression Analysis?* GeeksforGeeks.

<https://www.geeksforgeeks.org/what-is-regression-analysis/>