**PREDICTING AND COMPARING STUDENT PERFORMANCE IN ONLINE AND OFFLINE MODE USING MACHINE LEARNING TECHNIQUES**

**TEAM MEMBERS:**

1932042 - Samritha Shivani K

1932048 - Snekitha Sri D U

1932049 - Somasunder S

**ABSTRACT:**

Exam routine consists of arriving at your examination room, listening to someone discussing a subject you have no knowledge of, and finally taking your exam surrounded by 30 other nervous students. After the COVID-19 pandemic, this pattern has drastically changed, with classes being held online.Studying the effects of using the internet as a learning resource on students' performance is the focus of this project. The objective of this study was to evaluate and predict student performance in online mode and compare it to performance in offline mode to determine their learning potential.

**Keywords:** COVID-19, E-learning , Student performance , Machine Learning, Regression

**INTRODUCTION**

Over the past two years, the covid pandemic has caused a number of shifts globally. In particular, education has undergone a significant shift. This pandemic has seen an increased reliance on technology in education. In almost all educational institutions, online classes have become the norm.

There are pros and cons to this type of education. In online mode, students' mental and physical health do not remain stable, resulting in poor attention, interaction, concentration, an increase in screen time, and a lack of exposure. However, in the offline mode of education, students often interact with each other and encourage healthy competition, which helps them stay focused.

**Traditional teaching**

The traditional way of teaching refers to teaching where both teachers as well as students are involved directly in the education system. This system of education entails factors such as infrastructure and study material (textbooks, journals) that facilitate the teaching and learning process. Comparatively to distance learning, traditional learning is more organised and facilitates a free flow of communication and an open interaction between the teachers and students. The primary purpose of traditional education is to pass on skills, facts, and basic life standards which are considered critical for advancement in the material world by adults. Upon receiving this plan, students are expected to obediently accept and believe these agreed answers. Educators are the ones who transmit this knowledge and impose these behavioural standards.

**Remote Teaching**

During the days when distance learning wasn't a widely used mode of education, it was called a "correspondence course". The course content is delivered to the recipient through webinars, online streaming, live/recorded video lectures, a Learning Management System, or PowerPoint presentations, among other methods. Communication between students and teachers occurs via email, texts, and various other forms of messaging platforms.

**Challenges in remote teaching**

Education's latest innovation, eLearning, is making its mark despite presenting challenges to both teachers and students. Students need technical proficiency to decode course materials, and teachers need to put in thorough effort and time to structure the instruction material. Challenges are the difficulties and the obstacles faced by both students as well as teachers.

Some of the challenges in remote teaching are given below.

Ineffective Time Management

Studies indicate that students who master time management perform better academically and exhibit less anxiety. Moreover, that same study also found "many students have a difficult time striking a balance between their studies and daily lives" - and even worse, that ineffective time management was linked to poor sleep patterns and exhaustion. Distance learning can make time management especially challenging.

Lack of instant communication

In a classroom setting, communication happens instantly, allowing students to get answers and clarify confusion easily. Communication in an e-learning setting is often asynchronous, which means the teacher and student may have a gulf in communication.

Not Receiving Timely Feedback

A teacher's most meaningful way of engaging with a student is by providing feedback. If feedback is delayed by additional days or weeks due to an online format, students may be confused or unsure of your expectations, their progress, and their final grade in your class.

Not Receiving Clear Instructions or Expectations

Clarifying expectations for students is always crucial. Otherwise, they cannot be sure whether tasks and projects are being executed correctly. Establishing clear standards is a challenge in any classroom, but asynchronous communication can make it even more difficult.

**SOFTWARE PLATFORMS / TOOLS USED FOR EDUCATION**

As students had to stay at home due to the pandemic, teachers began creating online activities and lessons to keep kids engaged and focused. A variety of teaching tools are available to facilitate communication, including discussion boards, chat rooms, blog posts, and video conferencing. Most learning platforms, like Zoom and Google Classroom, and online platforms like Byjus, Nptel, Cousera, and YouTube Learning provide such an encouraging environment.

**NPTEL**

NPTEL is an acronym for National Programme on Technology Enhanced Learning which is an initiative by seven Indian Institutes of Technology (IIT Bombay, Delhi, Guwahati, Kanpur, Kharagpur, Madras and Roorkee) and Indian Institute of Science (IISc) for creating course contents in engineering and science. India needs many more teachers for effective implementation of higher education in professional courses. Therefore, methods for training young and inexperienced teachers to enable them carry out their academic responsibilities effectively are a must. NPTEL contents can be used as core curriculum content for training purposes.

A large number of students who are unable to attend scholarly institutions through NPTEL will have access to quality content from them.

**COURSERA**

Today, Coursera is a global online learning platform that offers anyone, anywhere, access to online courses and degrees from leading universities and companies.Coursera is an online learning platform founded by two Stanford University computer science professors. It offers thousands of online courses in partnership with over 200 of the world's leading universities and companies, including Yale, Princeton, UPenn, Google, IBM, Amazon, Facebook, and more.

The site offers individual courses as well as bachelor's and master's degree programs that reduce barriers to higher education. There are also professional certificate programs designed to aid workers in securing new roles or promotions.

**BYJU’S**

BYJU'S is an education tutoring app that runs on a freemium model, with free access to content limited for 15 days after registration.It was launched in August 2015, offering educational content for students from classes 4 to 12 and in 2019 an early learning program has started for classes 1 to 3.It also trains students for examinations in India such as IIT-JEE, NEET, CAT, IAS, and international examinations such as GRE and GMAT.

Academic subjects and concepts are explained with 12-20 minute digital animation videos.The biggest advantage of using Byju’s is the lower cost. If we compare the price of courses with metro cities, Byju’s courses are cheaper.

* It is cheaper than tuition classes.
* Students can learn anytime and anywhere.
* Students can learn with fun.
* Highly qualified teachers and professors.

**GOOGLE CLASSROOM**

Google Classroom is a free blended learning platform developed by Google for educational institutions that aims to simplify creating, distributing, and grading assignments. The primary purpose of Google Classroom is to streamline the process of sharing files between teachers and students.Google Classroom is a web-based learning environment. When logged into Google Classroom, students can collaborate with their peers and teachers. Google Classroom offers a digital safe space for students to view class announcements, access posted course content, view posted assignments, and turn in completed work.

**YOUTUBE**

Google has launched a new feature on YouTube to help students and educators who are stuck at home during lockdowns due to coronavirus outbreak. Dubbed as YouTube Learning, the new feature contains resources like curriculum-relevant topics in physics, math, and biology, language studies, study hacks, and much more, from YouTube's education-focused creators. Users can access YouTube Learning on mobile as well as on desktop using the Explore tab.

According to YouTube, in addition to curriculum-relevant topics, YouTube Learning also has video resources on topics related to photography and yoga. Google has said that its new feature will come in handy for people who are students and those who want to learn new skills.

**PROPOSED METHODOLOGY**

As part of our project, we used a manually collected dataset to perform the appropriate operations. As a result, this dataset contains almost 170 entries that are used for evaluating and predicting academic performance of students. The first step in the process is to clean data and preprocess it to eliminate imperfections. Next, we explore the data to find patterns that link each attribute together, and the final step is to feed the data into machine learning models so that we can get a decent fit and an accurate prediction.

**1.Data collection and feature exploration**

Data acquisition:

Google forms were created and distributed to students of the Computing Department at Coimbatore Institute of Technology. Nearly 170 entries were entered into the form. Aside from name, department, age, gender, and the CGPA for the first four semesters, the forms asks for about 19 questions related to the students' academic performance before and after COVID and a lot of other questions.

Some of the questions are given below.

**Q1.**Before COVID-19: How much time do you spend using the digital tools in learning (in hrs) ?

**Q2.**After COVID-19: How much time do you spend using the digital tools in learning (in hrs) ?

**Q3.**Before COVID-19: I always use digital tools (mobile, laptop, i-pad) in studying.

**Q4.**After COVID-19: I always use digital tools (mobile, laptop, i-pad) in studying.

**Q5.**Before COVID-19: I have fixed hours for bedtime and wake-up.

**Q6.**After COVID-19: I have fixed hours for bedtime and wake-up.

**Q7.**Continuous exposure to electronic screens in online learning is tiring and exhausting

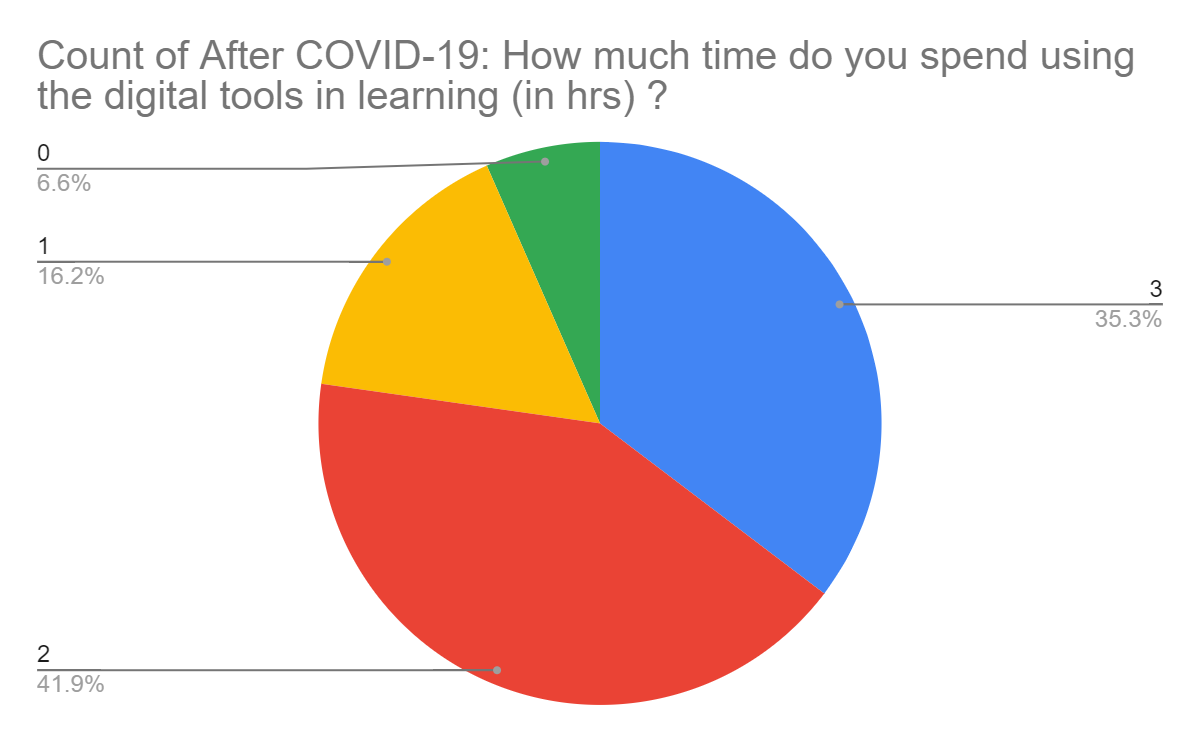
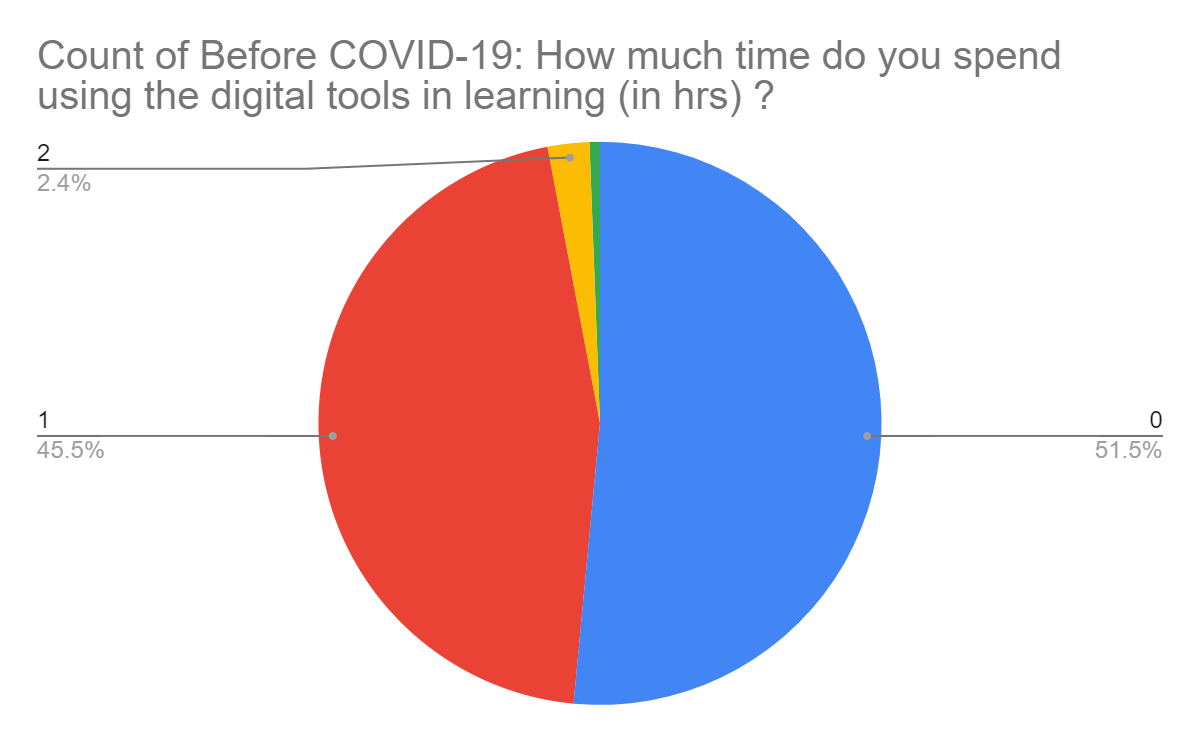
**Q8.**Prolonged use of e-learning tools often leads to boredom, nervousness, and tension

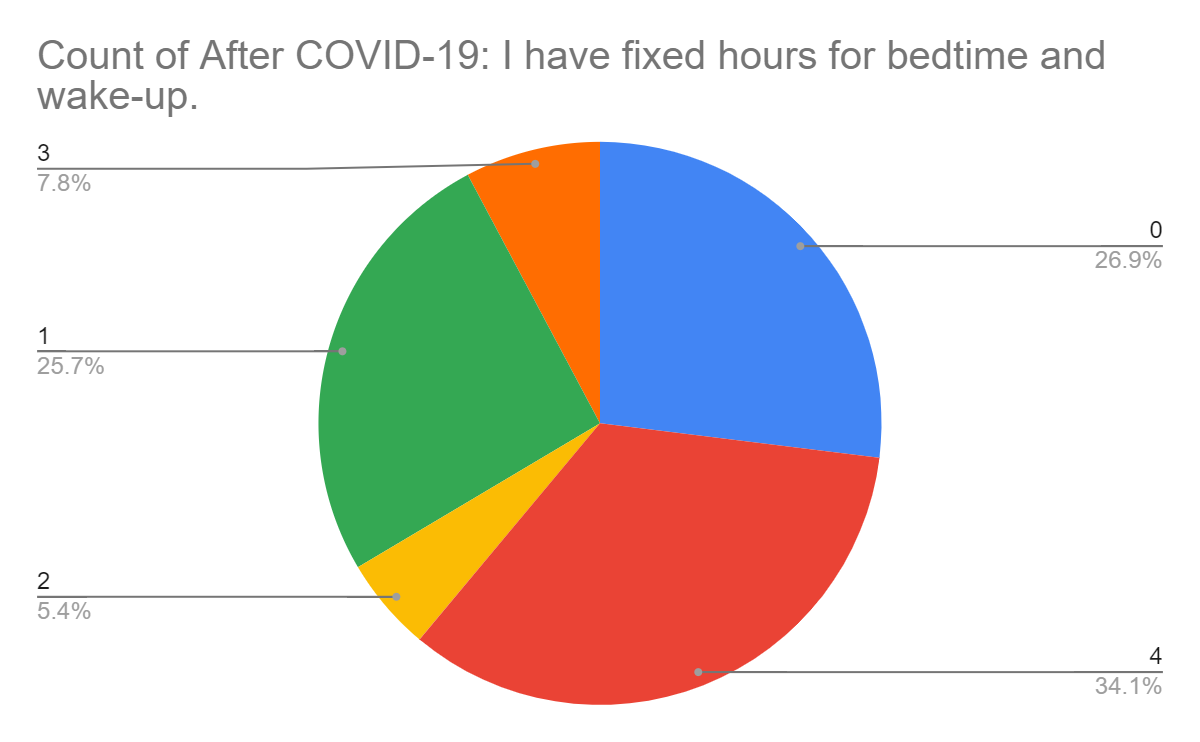
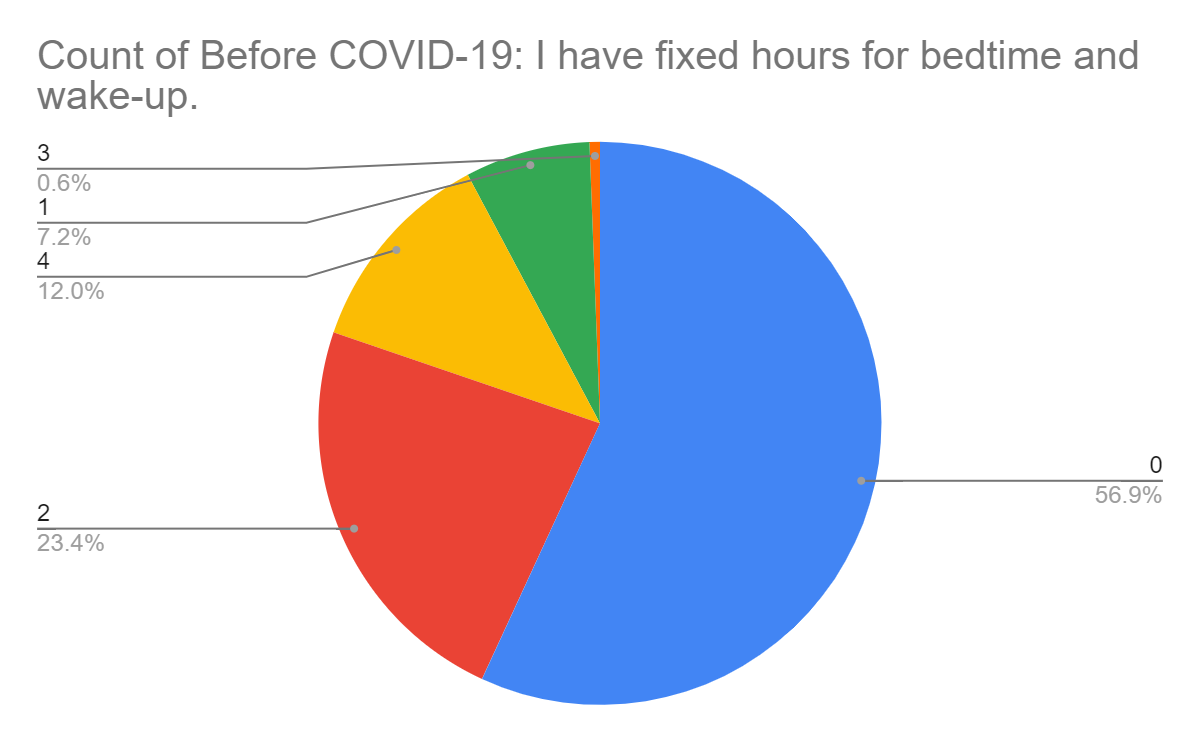
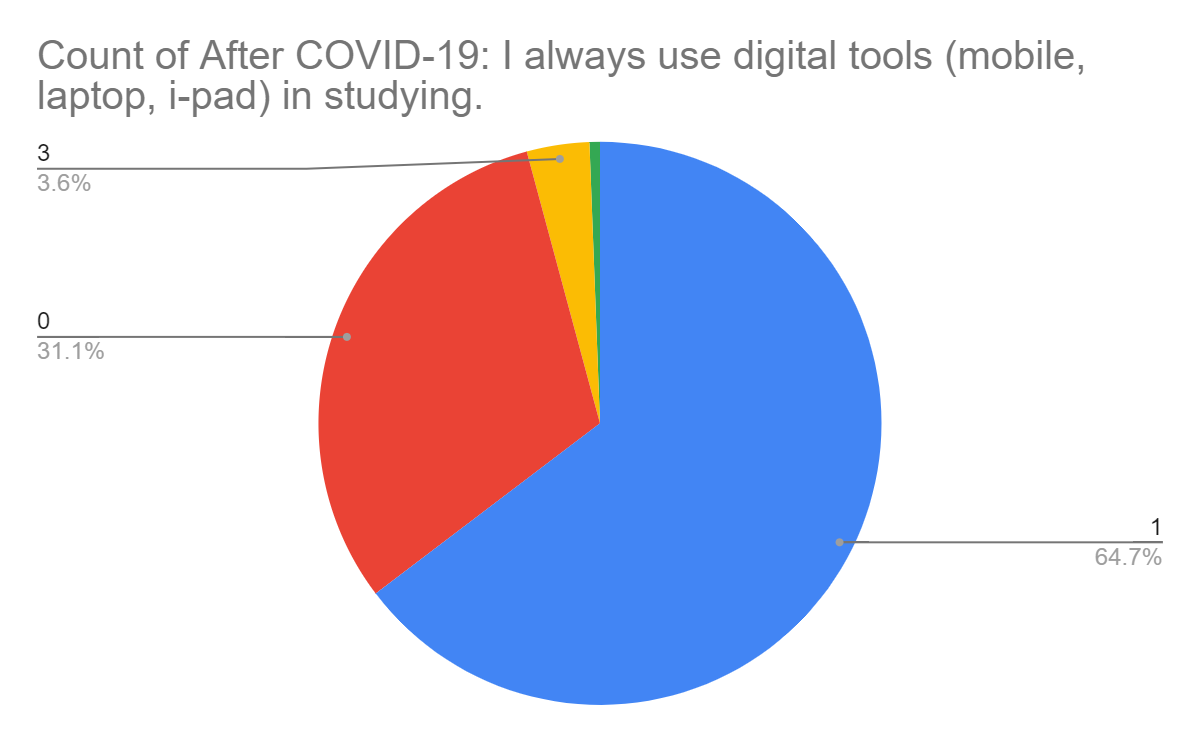
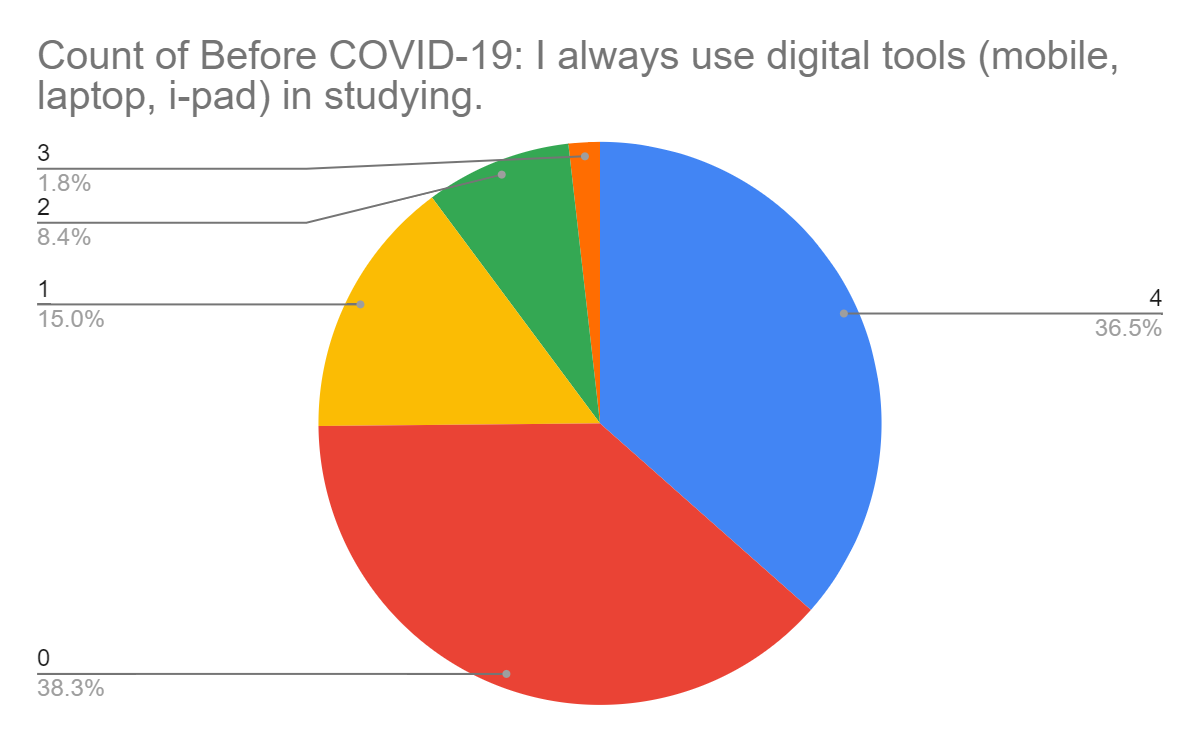
**Q9.**I don’t recommend continuing with the online learning model because it is socially and psychologically unhealthy.

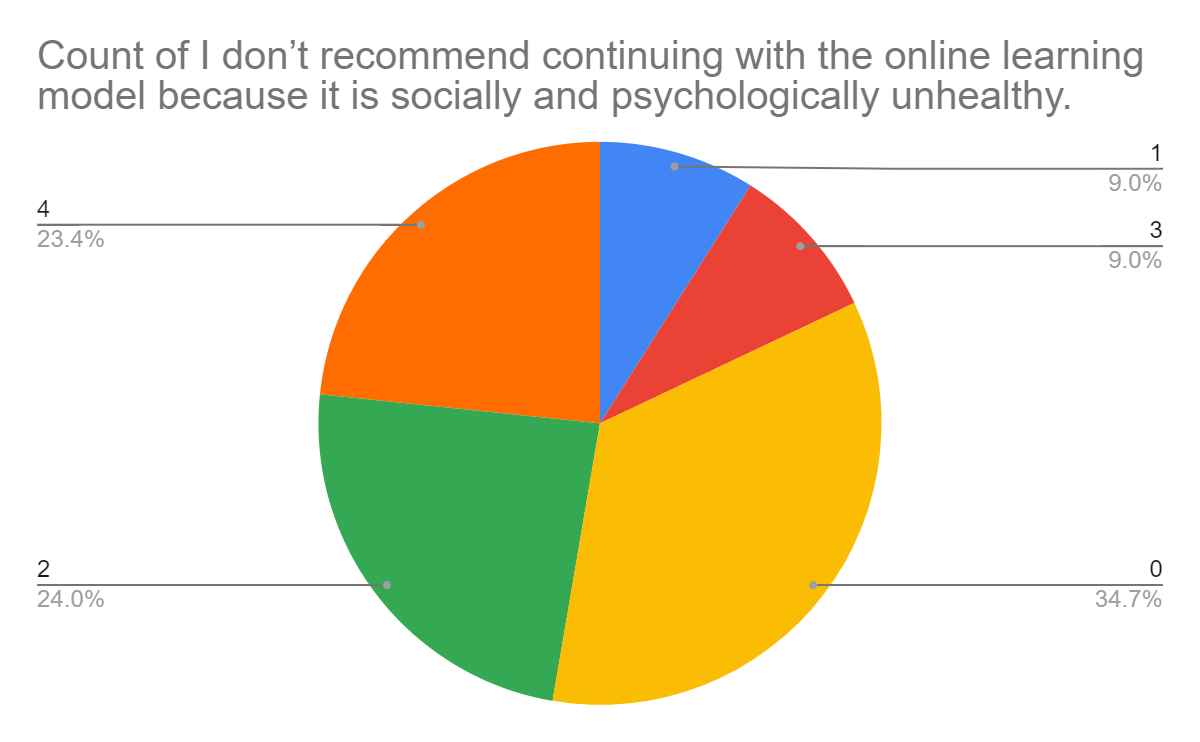
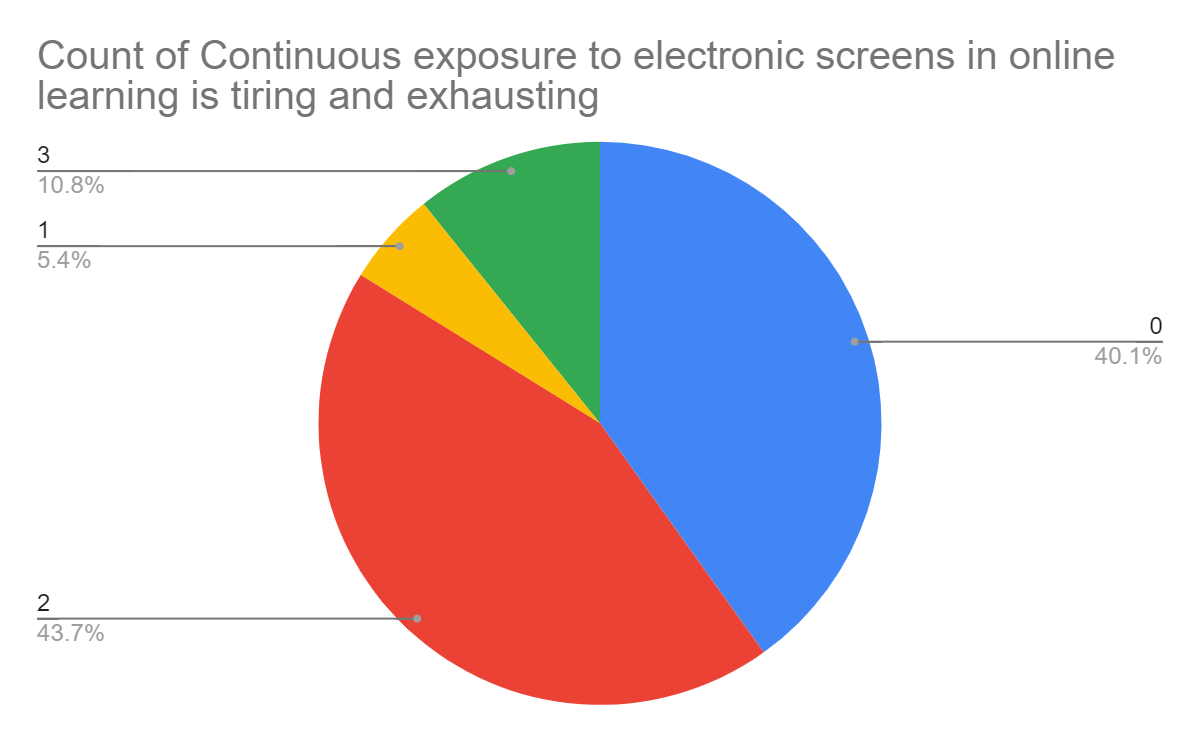
**Q10.**The volume of assignments via e-learning led to confusion, frustration and poor performance

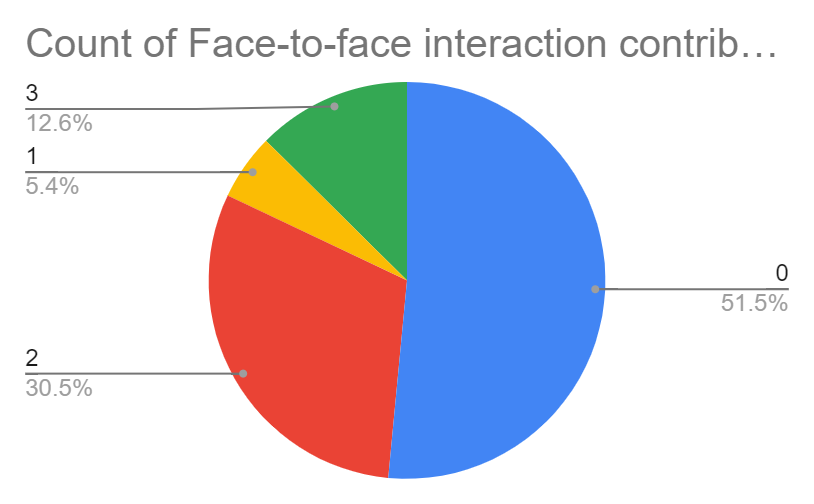
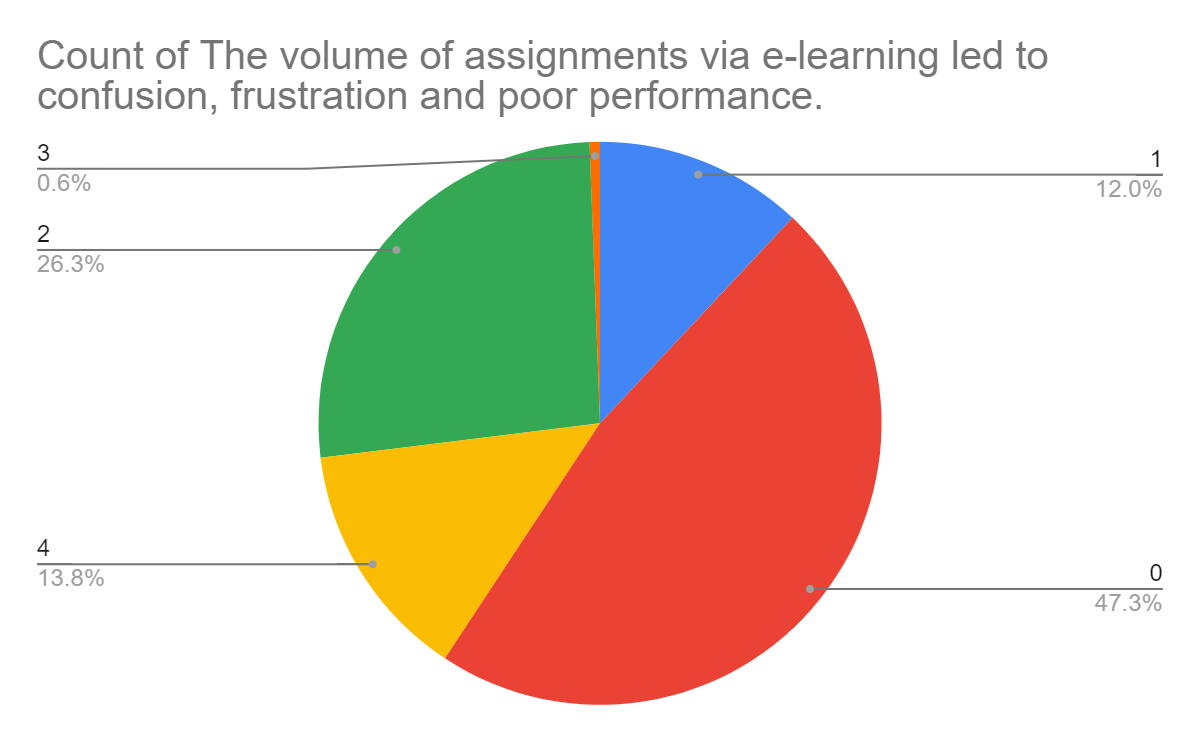
**Q11.**Face-to-face interaction contributes significantly to boosting students’ academic achievement.

| **Q1** | **Q2** | **Q3** | **Q4** | **Q5** | **Q6** | **Q7** | **Q8** | **Q9** | **Q10** | **Q11** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 3 | 4 | 1 | 0 | 0 | 0 | 4 | 1 | 1 | 0 |
| 1 | 2 | 0 | 1 | 0 | 0 | 0 | 3 | 3 | 0 | 0 |
| 1 | 2 | 4 | 1 | 0 | 4 | 2 | 0 | 0 | 4 | 0 |
| 1 | 3 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1 | 3 | 4 | 0 | 0 | 4 | 0 | 0 | 0 | 4 | 0 |
| 1 | 3 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 2 | 0 | 0 | 2 | 3 | 2 | 2 | 2 | 2 | 0 |
| 0 | 2 | 4 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 | 0 |
| 0 | 2 | 4 | 1 | 0 | 4 | 0 | 0 | 4 | 4 | 3 |
| 1 | 3 | 4 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

****

****

****

****

Google form link for,

SURVEY 1:

<https://docs.google.com/forms/d/1XQtowAUW0sGzrCy-0aySlnbZ4NyBPZLulukmy27Ljz0/edit>

SURVEY 2:

<https://docs.google.com/forms/d/1CU1WgsxT8T_lOSQh0-cXp4CO5ConJ9m9GqITTGAS9zs/edit>

**2.Data Preprocessing**

Preprocessing of data includes cleaning the data and converting it into machine readable form. First, the binning method is used for grouping related values together in bins to reduce the number of distinct values. Then the data is checked for any missing values which is followed by conversion of decimal scores into integer values for model fitting purposes. And finally, the selected features that are in categorical form are converted to numeric form using label encoding.

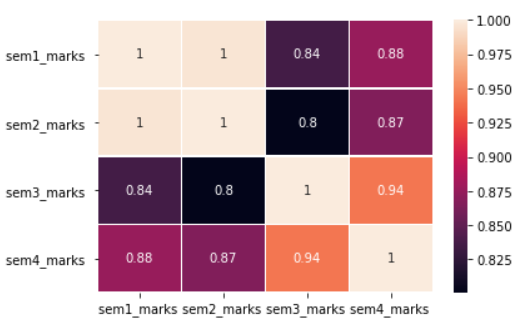
**3.Feature Selection:**

1. Univariate Selection:

All the semester marks and the questionnaires are unique in nature and it is highly necessary for the prediction of output. We have removed the Name and Roll number of students.

1. Correlation Matrix with Heatmap :

We can clearly see the correlation between the semester marks, which are very fairly correlated. There is not much difference between sem1 and sem2 marks (Offline classes) , similarly sem3 and sem4 marks (Online classes) are similar to each other. But there are huge variations between the offline and online exams. We can be used for better results.

****

**4.Machine learning models - Regression**

Multiple linear regression

In regression models, a line is fitted to the data to describe the relationship between variables. A regression equation calculates the impact of changing independent variables on a dependent variable.

When two or more independent variables are combined with a single dependent variable, multiple linear regression is used to estimate their relationship. MLR is an extension of Linear Regression but has more than one independent variable. MLR indicates the degree to which two or more independent variables have a relationship with one dependent variable. It is an estimate of the dependent variable at a particular value of the independent variable .

Gradient Boosting Regressor

Gradient Boosting Regression is a Machine Learning technique which is used for both classification and regression problems.Combined weak learners or weak predictive models are combined using the gradient boosting algorithm to create an ensemble model. For fitting the model for predicting continuous values, this model algorithm is used.

By using multiple decision trees of fixed size as weak learners and weak predictive models, an additive mode is created. As a starting point, the mean value of the target values is used as a constant. A second stage involves fitting decision trees or estimators to predict the negative gradients for each sample.

Extra Trees Regressor

Extra Trees Regressor works the same way as Gradient Boosting Regressor by using several decision trees to make predictions. Using the training dataset, Extra Trees creates a large number of trees without pruning them. As a rule of thumb, regression predictions are averaged by using decision trees and classification predictions are determined by maximum voting.

This algorithm is also called Extremely Randomized Trees. Here to introduce a lot of variation, the trees are built in a unique way. There exists extreme randomness in the chosen features and the splits and this algorithm shows a low variance.

Random Forest Regressor

Random Forest algorithm is a technique in supervised learning which is used for both regression and classification problems. It is also an ensemble technique where a lot of models are combined to make accurate predictions. The training set is randomly sampled and it involves a lot of trees hence the name random forest.

This algorithm combines several decision trees to construct a model and outputs the average prediction of each tree. Each individual tree chooses a random sample from the training set during the split operation which helps with avoiding overfitting of the model. So a potential use of every feature that exists in the dataset makes sure that this ensemble model does not depend on a particular predictive feature.

**EXPERIMENTAL RESULTS**

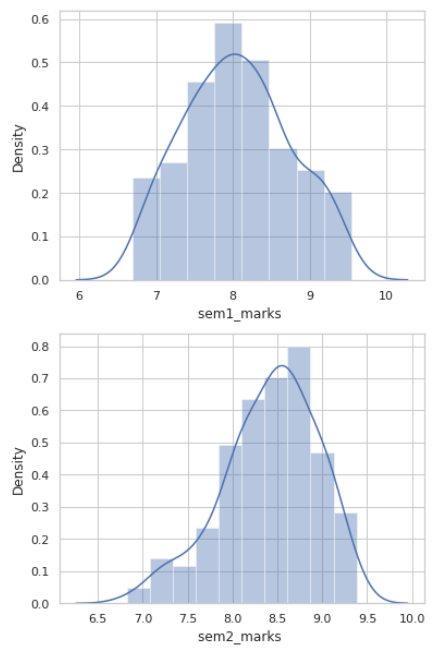
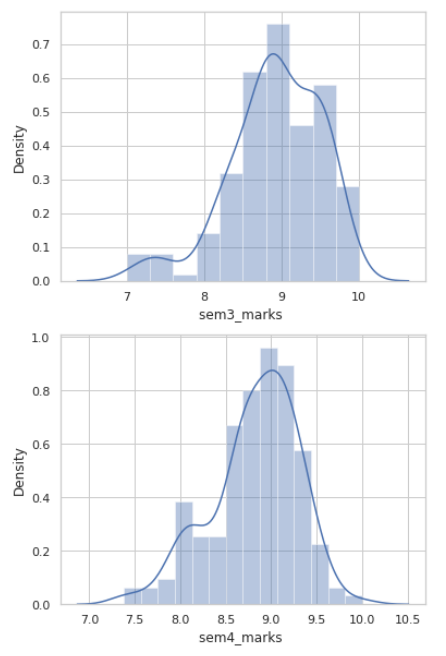
Four regression models were used to make the prediction and the results are given below.

| **MODEL** | **ACCURACY** |
| --- | --- |
| Multiple Linear Regression | 61.99% |
| Gradient Boosting Regressor | 56.38% |
| Extra Trees Regressor | 70.03% |
| Random Forest Regressor | 98.10% |

Comparing each model used on this dataset, Random Forest Regressor has the maximum accuracy of 98.10%.

**INFERENCE**

A rationale for REGRESSOR instead of other methods or models: In this context, we are delving into marks and the factors that influence them. It is true that there is a relationship, but it is numerical. This is why only the regression model is considered to be the correct choice.No matter what classifier models we use, we won't get an accuracy higher than 10-20%, but we can get a minimum of 50% accuracy with regression. This is because there is an empirical relationship between them.

** **

From above graphs of offline and online exams, we can easily give interference that :

Semester 1 scores are normally distributed and semester 2 score is also to some extent normally distributed, whereas semester 3 is not normally distributed and semester 4 is also not normally distributed but it is comparatively better than semester 3.

Therefore, there is a lot of randomness in semester 3 then following semester 4 and 2.

Randomness in Scores :

**max randomness > Semester 3 >semester 4 > semester 2 > Semester 1 > Min Randomness**

Online exams have a lot of randomness , whereas offline exams give a stable score.

Possible reasons :

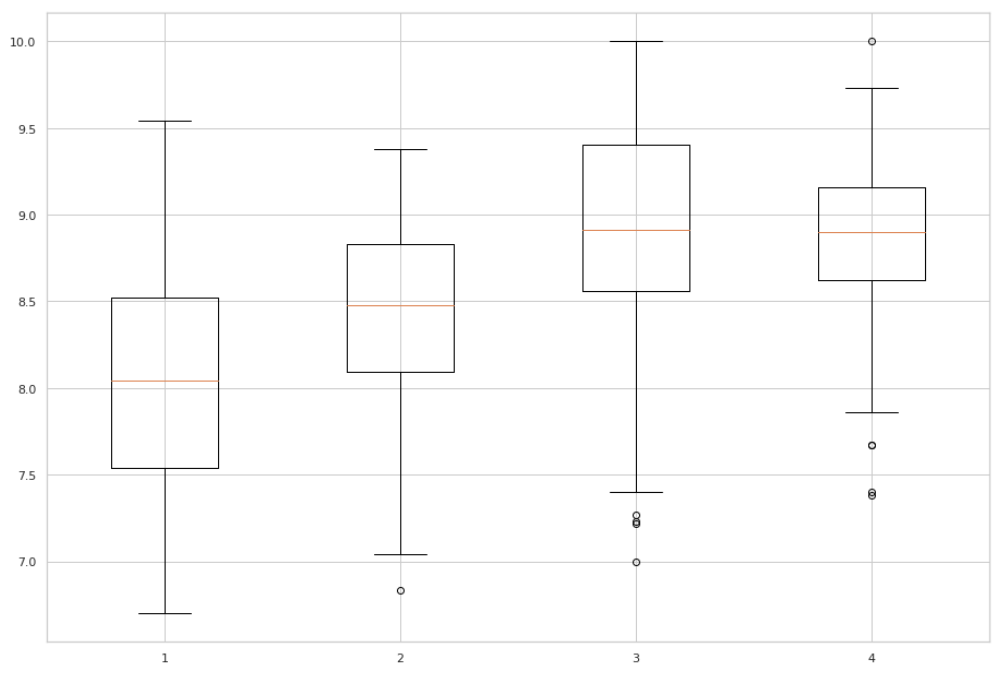
1. Cheating in online exams

2. improper understanding or contact between student and staff

3. Uneven markings

4. Online Gradings

**Outlier Detection:**

****

By using box plots we can say that there exist more outliers in online exams than offline exams.

No. of Outliers in Scores :

**Max no. of Outliers > Sem 3 > Sem 4 > Sem 2 > Sem 1 > Min no. of Outliers**

**CONCLUSION**

This project is devoted to the development of a student performance evaluation system. Despite the fact that online classes offer several benefits such as flexibility, accurate grades per student and cost benefits, it does not provide a healthy environment for students to learn and compete. Therefore, offline classes are comparatively better for people and the environment since online classes make students' minds idle and inactive.