SILIGURI INSTITUTE OF TECHNOLOGY

PROJ- CS881

E-LEARNING AS A SUSTAINABLE DIGITAL TOOL

BY

IT_PROJ_2024_08

Name of Students	Roll No.
1. Shubham Divyanshu	11900220008
2. Ayush	11900220017
3. Atul Krishna Gupta	11900220019
4. Priyatosh Nandy	11900220021

Under the Guidance

of

Prof. Sathi Ball

Submitted to the Department of **Information Technology** in partial fulfillment of the requirements for the award of the degree Bachelor of Technology in **Information Technology.**

Year of Submission: 2024

Siliguri Institute of Technology

P.O. SUKNA, SILIGURI, DIST. DARJEELING, PIN:734009

Tel: (0353)2778002/04, Fax: (0353) 2778003

DECLARATION

-

This is to certify that Report entitled "**E-LEARNING AS A SUSTAINABLE DIGITAL TOOL**" which is submitted by me in partial fulfillment of the requirement for the award of degree B.Tech. in **Information Technology** at **Siliguri Institute of Technology** under **Maulana Abul Kalam Azad University of Technology**, West Bengal. We took the help of other materials in our dissertation which have been properly acknowledged. This report has not been submitted to any other Institute for the award of any other degree.

Date:

SN	Name of the Student	Roll No	Signature
1	Shubham Divyanshu	11900220008	
2	Ayush	11900220017	
3	Atul Krishna Gupta	11900220019	
4	Priyatosh Nandy	11900220021	

CERTIFICATE

This is to certify that the project report entitled

E-LEARNING AS A SUSTAINABLE DIGITAL TOOL

submitted to **Department of Information Technology of Siliguri Institute of Technology** in partial fulfilment of the requirement for the award of the degree of **Bachelor of Technology in Information Technology** during the academic year **2023-24,**is a bonafide record of the project work carried out by them under my guidance and supervision.

Project Group Number : 08			
SN	**Name of the students**	**Registration No**	**Roll No**
1.	**Shubham Divyanshu**	**201190100210032**	**11900220008**
2.	**Ayush**	**201190100210023**	**11900220017**
3.	**Atul Krishna Gupta**	**201190100210021**	**11900220019**
4.	**Priyatosh Nandy**	**201190100210019**	**11900220021**

Signature of Project Guide

Name of the Guide:

Signature of the HOD

Department of Information Technology

Acknowledgement

We would like to express our sincere gratitude to our project advisor, Mrs. Sathi Ball, for her invaluable guidance and support throughout this project. We are also thankful to **Siliguri Institute of Technology** for providing the necessary resources and environment for this research. Special thanks to our team members, for their constructive feedback and invaluable inputs, and to our colleagues for their unwavering support. Lastly, we would appreciate the participants in the study and the pioneers in e-learning whose work has inspired this research. Thank you all for your contributions to the successful completion of this project, **"E-LEARNING AS A SUSTAINABLE DIGITAL TOOL."**

Signature of all the group members with date

1.

2.

3.

4.

Table of Contents

Abstract

The project "E-learning as a Sustainable Digital Tool" aims to explore and harness the potential of digital technologies in the realm of education to foster sustainability. In an era marked by rapid technological advancements, the integration of e-learning platforms has become increasingly prevalent in educational settings. This project seeks to investigate how e-learning can contribute to sustainable practices in education, encompassing environmental, social, and economic dimensions. To achieve these objectives, the project will employ a multi-disciplinary approach, integrating insights from education, environmental science, social sciences, and economics. Data will be collected through surveys, case studies, and interviews with educators, students, and industry experts. The findings of this research will provide valuable insights into the potential benefits and challenges of implementing e-learning as a sustainable digital tool.

Objectives

1. Understand and identify the factors (construct/variable) affecting the adoption of the e-learning process.

2. Using that factors create a set of questioners.

3. Explore all the methods used to investigate the adoption of the e-Learning system and measure effectiveness for its long-term sustainability.

Introduction

In the ever-evolving landscape of education, the integration of technology has emerged as a transformative force, bringing about significant changes in the way knowledge is disseminated and acquired. One such paradigm shift is the advent of e-learning, a sustainable digital tool that has revolutionized the educational ecosystem.

E-learning, short for electronic learning, harnesses the power of digital resources and the internet to facilitate educational experiences beyond traditional classrooms. It has seamlessly woven itself into the fabric of learning institutions, corporate training programs, and lifelong education initiatives, offering a sustainable and flexible alternative to conventional methods. This transformative tool not only addresses the immediate challenges posed by global events, such as the COVID-19 pandemic, but also aligns with broader sustainability goals, making education more accessible, inclusive, and environmentally friendly.

The sustainability of e-learning is evident on various fronts. Firstly, it significantly reduces the carbon footprint associated with traditional education. The digital delivery of courses eliminates the need for printed materials, commuting to physical locations, and the energy consumption associated with brick-and-mortar facilities. Secondly, e-learning promotes inclusivity by breaking down geographical barriers. Learners from diverse backgrounds, irrespective of their location, can access high-quality educational content with just an internet connection. This democratization of education ensures that knowledge becomes a shared global resource, transcending socio-economic constraints and fostering a more equitable distribution of learning opportunities.

Furthermore, the adaptability and scalability of e-learning make it a sustainable solution for both formal and informal education. It is clear that this digital tool is not just a temporary fix but a transformative force with lasting implications. Its sustainability lies not only in its immediate adaptability to unforeseen challenges but also in its long-term impact on shaping a more accessible, inclusive, and environmentally conscious educational landscape. In the chapters that follow, we delve deeper into the various facets of e-learning that contribute to its sustainability and explore how it continues to redefine the contours of education in the 21st century.

System Analysis

i) Identification of Need

The need for adoption of e-learning processes are due to the following factors:-

a) **Metacognity Activity:** Higher metacognitive activity can positively impact e-learning adoption by enhancing students' ability to plan, monitor, and evaluate their learning strategies.
b) **Performance:** E-learning platforms that provide measurable and immediate feedback can contribute to a sense of accomplishment, motivating learners to continue using the platform.
c) **Reliability:** The reliability of e-learning platforms, including consistent access to content and functionalities, is crucial. Unreliable systems can lead to frustration and may deter users from fully embracing e-learning.
d) **Social Presence and Media Synchronicity:** E-learning platforms facilitate social interaction through live chats and provide synchronous communication tools to enhance overall learning experience.

ii) Preliminary Investigation

We use education data mining (EDM) methods to extract patterns and drawing interfaces from large and complex datasets. In our project, we are using it to predict the reliability and performance of a student based on their activities such time spent on e-learning courses before and after covid times, devices used and many more.

iii) Feasibility Study (Software Requirement Specification)

Data Requirements

- **Student Performance Data:** Grades in gpa.
- **eLearning Interaction Data**: Time spent on platform, number of logins, activity completion.
- **Demographic Data:** Age, gender, level/year, etc.
- **Pre-processing:** Ensure data is cleaned, normalized, and anonymized.

Technical Requirements:

- **Hardware:** Windows 10 intel i3 11th gen CPU with 120Hz refresh rate.
- **Software:** Jupyter Notebook and few python libraries and frameworks (e.g.numpy, pandas, matplotlib, seaborn, scikit-learn, TensorFlow, category_encoders).
- **Data Storage:**Use of excel workbook to store dataset of nearly 800 students.

Skill Requirements:

- **Expertise in ML:** Team members should have knowledge of machine learning algorithms, data preprocessing, and model evaluation.
- **Domain Knowledge:** Understanding of educational metrics and eLearning platforms.

iv) Project Planning and Scheduling:

a) **Dataset:** We collected a dataset of about 800 students containing information such as demographic details like age, year of study, gender, and GPA.This dataset is valuable because it captures students' use of digital tools for studying and considers the psychological impact of their excessive use, which is crucial for understanding academic performance.

1	Gender	Level/Year	Age	Your cumulative average (GPA)	Before COVID-19: Which of the following digital tools
2	Female	Second/ Sophomore	18-24	80-89 / 3-3.49	Mobile phone
3	Male	Other	+30	+90 / +3.5	Laptop
4	Female	First/Freshman	18-24	+90 / +3.5	Other

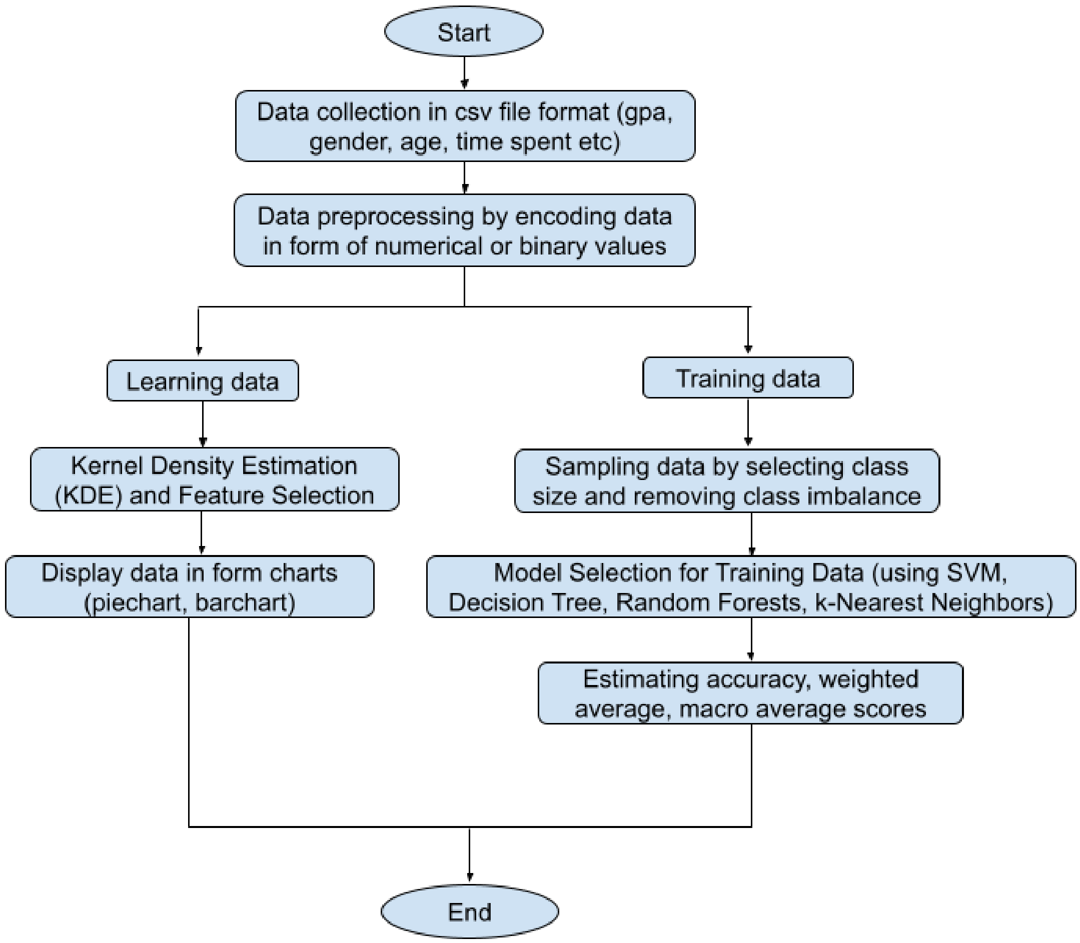
b) **Data Preparation:**

1. Converting all Likert responses into numerical values (‘Strongly Disagree’:0, ‘Disagree’:1, ‘Strongly Agree’:4,‘Uncertain’:2, ‘Agree’:3)
2. Performing One-Hot Encoding for all categorical features.
3. And label encoding for our target value (GPA : ‘Below 60 / Below 2.0’:0, ’60–69 / 2–2.49':1, ‘70–79 / 2.5–299’:2, ‘80–89 / 3–3.49’:3, ’+90 / +3.5':4) and Hour values (‘3–6’:1, ‘6–9’:2, ‘9–12’:3, ‘+12’:4, ‘1–3’:0).
4. As there are only 4 rows with GPA below 2.0, we remove them.
5. After removing all the NaN values we are left with 781 rows and 46 columns. Finally, we have our cleaned data set as shown below:

	Your cumulative average (GPA)	Before COVID-19: How much time do you spend using the digital tools in learning?	After COVID-19: How much time do you spend using the digital tools in learning?	Before COVID-19: I always use digital tools (mobile, laptop, i-pad) in studying.	After COVID-19: I always use digital tools (mobile, laptop, i-pad) in studying.	Before COVID-19: When I use the mobile phone, tablet or laptop in e-learning, I cannot concentrate and I am distracted.
0	3	2	3	2	3	4
1	4	0	1	3	4	1
2	4	0	1	1	4	4
3	2	2	0	2	4	3
4	2	0	0	3	3	1

v) Control Flow Diagram

The following flowchart depicts the plan to be followed as per SRS document:



System Design

i) Modularisation details: The project is basically subdivided into two modules or data sets- **learning data set** which is used for Kernel Density (KDE) Estimation and Feature Selection and to obtain graphs and curves necessary to study the impact of e-learning platforms on the performance of the students, and

Training data set which is used to find the reliability scores of elearning courses through various regression models such as SVM, Decision Tree, Random Forests, k-Nearest Neighbors.

ii) Data integrity and constraints

a) **Feature Engineering:** Create new features or transform existing ones to improve model performance and ensure data integrity. This includes techniques like one-hot encoding, feature scaling, and dimensionality reduction.

   category_encoders package in python helps in performing this task.

b) **Imbalanced Classes:** If our dataset has imbalanced classes, employ techniques such as oversampling, undersampling, or using algorithms that are inherently robust to class imbalance to maintain data integrity and prevent bias.

   Imblearn package in python is used in this case.

iii) Database Design

a) **Data Schema Design:** Designing an appropriate database schema that accommodates the data needed for the e-learning platform. This includes tables for users, courses, lessons, quizzes, user progress, etc.

b) **Normalization:** Ensuring the database is normalized to reduce redundancy and improve data integrity. This involves organizing the database structure into well-defined tables and relationships to minimize data duplication.

c) **Scalability:** Designing the database to handle potential growth in data volume and user base over time. This may involve choosing scalable database technologies and architectures.

Coding and Testing

i) Importing Libraries

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
from tensorflow import keras
from tensorflow.keras import layers
import category_encoders as ce
```

ii) Uploading Dataset

```
viewData = pd.read_csv('elearn.csv')
viewData
# Dropping null values
viewData=viewData.dropna(axis=0)
```

iii) Encoding Values

```
# showing target values
viewData['Your cumulative average (GPA)'].unique()

# OrdinalEncoder function maps every string data to its equivalent numeric data for easy
analysis
encoder = ce.OrdinalEncoder(cols=['Your cumulative average (GPA)'],return_df=True,
                    mapping=[{'col':'Your cumulative average (GPA)',
'mapping':{'Below 60 / Below 2.0':0,'60-69 / 2-2.49':1,'60-69 / 2-2.9':1, '70-79 / 2.5-299':2,
'80-89 / 3-3.49':3,'+90 / +3.5':4, '60-69':1, '80-89':3, '+90':4,
    '70-79':2, 'Below 60':0}}])

# replaces the data with encoded values
viewData = encoder.fit_transform(viewData)
viewData['Gender'].unique()

# deleting column Gender and creating it a new column 'Male or not'
viewData['Male'] = viewData['Gender'].map({'Male':1, 'Female':0})
viewData = viewData.drop(['Gender'], axis=1)
viewData['Level/Year'].unique()

# correcting field values of 'Level/Year'
roles={'Fourth':'Fourth/Senior', 'Third':'Third/Junior', 'Second':'Second/ Sophomore',
'First':'First/Freshman'}
viewData = viewData.replace(roles)
# creating dummy fields/columns of above
dummy = pd.get_dummies(data=viewData['Level/Year'])
```

```
# removing duplicate values and deleting the original column
viewData = pd.concat([viewData, dummy], axis=1)
viewData = viewData.drop(['Level/Year'], axis=1)
viewData['Age'].unique()

# creating dummy fields of column 'Age' and deleting the original column
ageDummy = pd.get_dummies(data=viewData['Age'])
viewData = pd.concat([viewData, ageDummy], axis=1)
viewData = viewData.drop(['Age'], axis=1)
```

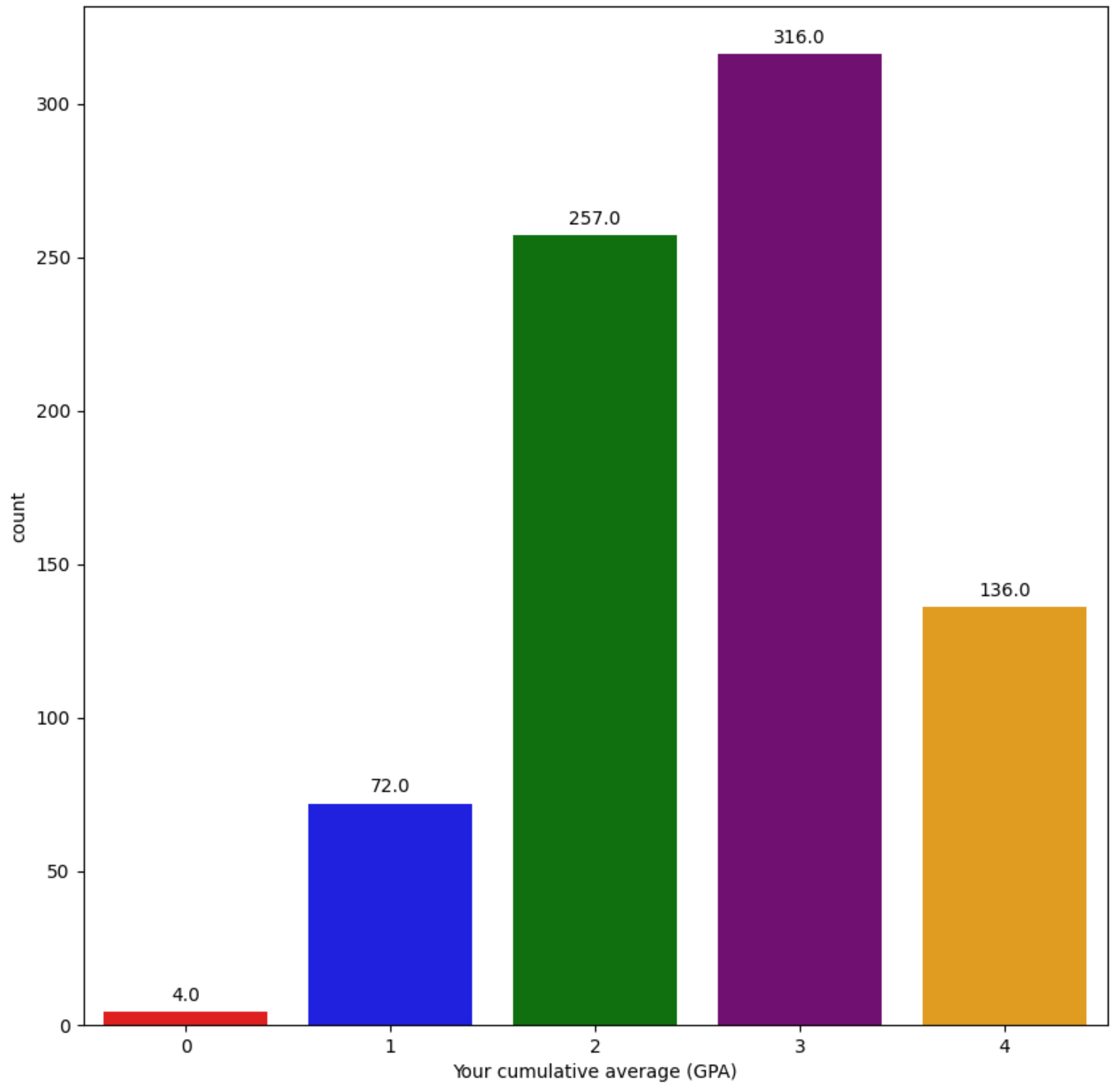
iv) Data Analysis for plotting graphs

```
# code for pie chart depicting time spent in digital tools in e-learning before covid-19
plt.figure(figsize=(8, 8))
grouped_data = viewData['Before COVID-19: How much time do you spend using the digital tools in learning?'].value_counts()
grouped_data.plot(kind='pie', autopct='%1.1f%%', startangle=90, title='Before COVID-19: Time spent using digital tools in learning')
plt.axis('equal')
plt.legend(labels=grouped_data.index, loc='best')
plt.show()

# code for pie chart depicting time spent in digital tools in e-learning after  covid-19
plt.figure(figsize=(8, 8))
grouped_data = viewData['After COVID-19: How much time do you spend using the digital tools in learning?'].value_counts()
grouped_data.plot(kind='pie', autopct='%1.1f%%', startangle=90, title='After COVID-19: Time spent using digital tools in learning')
plt.axis('equal')
plt.legend(labels=grouped_data.index, loc='best')
plt.show()
```

v) Kernel Density Estimation

```
# code for comparison between KDE plots on time spent on digital tools for learning (before & after Covid) v/s CGPA
testview = viewData
fig = plt.figure(figsize=[15,5])
plt.tight_layout()
for i in range(2):
fig.add_subplot(1,  2, i+1)
sns.kdeplot(data=testview,x=testview.columns[i+1],hue='Your cumulative average (GPA)')
  if i == 16:
plt.xlim([-50,300])
sns.despine()
```

```
# code for barchart depicting no. of students vs their cumulative GPA)
custom_palette = sns.color_palette(['red', 'blue', 'green', 'purple', 'orange'])
plt.figure(figsize=(10, 10))
gplot = sns.countplot(data=testview, x='Your cumulative average (GPA)', hue='Your
cumulative average (GPA)', palette=custom_palette, legend=False)
for p in gplot.patches:
gplot.annotate(format(p.get_height(), '.1f'),
               (p.get_x() + p.get_width() / 2., p.get_height()),
               ha='center', va='center',
xytext=(0, 9),
textcoords='offset points')
plt.show()
```



vi) Testing and Training data (using SMOTE)

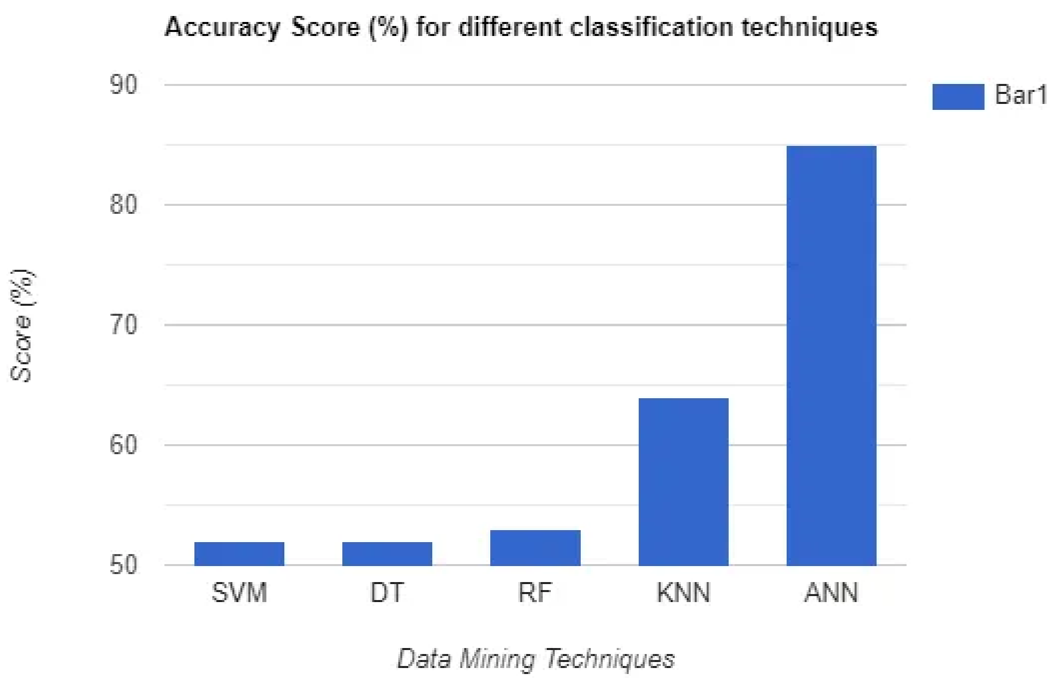
```
# Dividing the datset into parameters columns (X) and target value (y)
X=testData.iloc[:, 1:]
y=testData.iloc[:, [0]]

# Applying SMOTE to get rid of class imbalance
from imblearn.over_sampling import SMOTE
oversample = SMOTE()
X, y = oversample.fit_resample(X, y) # X is your feature matrix and y is your target variable
```

NB: Synthetic minority oversampling technique (SMOTE) generates new rows for the minority class that is under-represented in our dataset.

vii) Regression testing using classification models

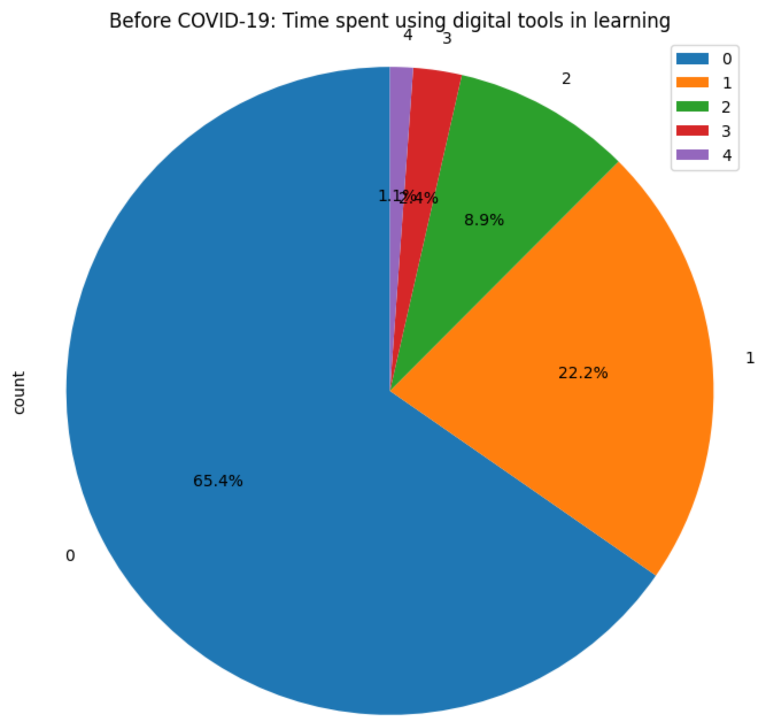
```
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.multiclass import OneVsRestClassifier
from sklearn.multiclass import OneVsOneClassifier

# Train test split size is taken 0.33
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42,
shuffle=True)

# Predicting score through Support Vector Machine (SVM)
from sklearn.svm import SVC
model = SVC()
ovo = OneVsOneClassifier(model)
ovo.fit(X_train, y_train.values.ravel())
prd=ovo.predict(X_test)
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(prd, y_test))
print(classification_report(prd,y_test))

# Predicting score through Decision Tree
from sklearn.tree import DecisionTreeClassifier
dtree=DecisionTreeClassifier()
ovo = OneVsOneClassifier(model)
ovo.fit(X_train, y_train.values.ravel())
pred=ovo.predict(X_test)
print(classification_report(y_test, pred))
```

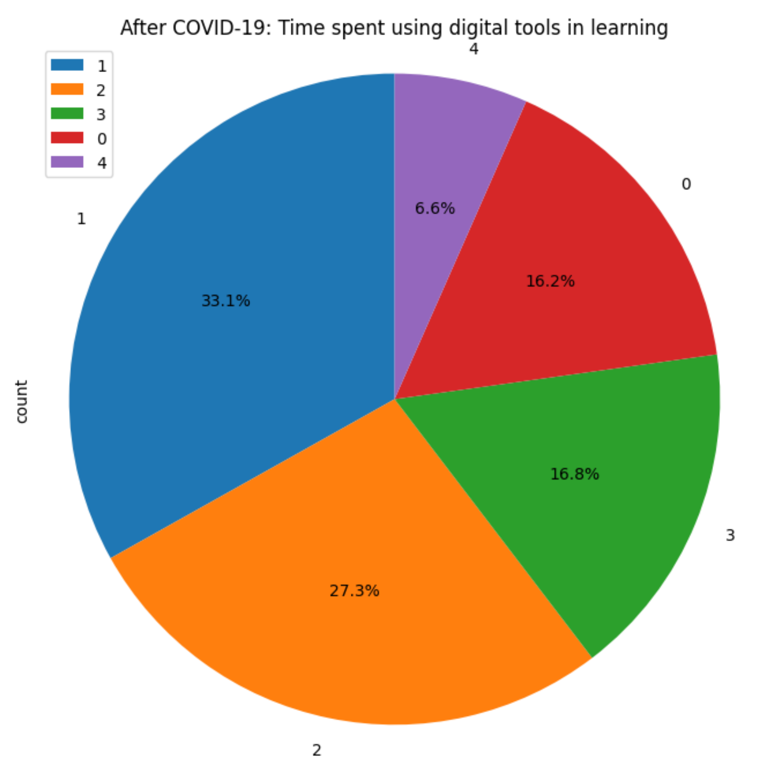
```
#Predicting score through Random Forest
from sklearn.ensemble import RandomForestClassifier
rd=RandomForestClassifier(n_estimators=200) # Can change value
ovo = OneVsOneClassifier(model)
ovo.fit(X_train, y_train.values.ravel())
rd_pred=ovo.predict(X_test)
print(classification_report(rd_pred, y_test))

# Predicting score through K-Nearest Neighbors
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=15)
ovo = OneVsOneClassifier(neigh)
ovo.fit(X_train, y_train.values.ravel())
y_pred=ovo.predict(X_test)
print(classification_report(y_pred, y_test))
```



Project Score and Cost Estimation

i) Data Analysis using charts

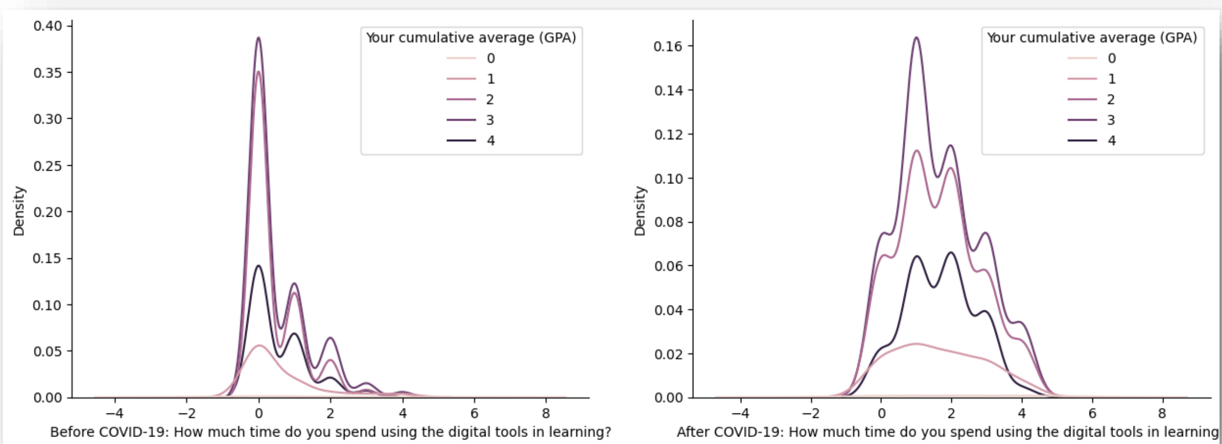




Work hours	Value assigned	Before covid (%)	After covid (%)
1-3	0	65.4	16.2
3-6	1	22.2	33.1
6-9	2	8.9	27.3
9-12	3	2.4	16.8
+12	4	1.1	6.6

From the following two images queried from the dataset we find that the percentage of students using digital tools for more than 3–6 hours increased by 10.9% while those using it for more than 9–12 hours increased by 14.4%. From this we are able to clearly visualize the increase in device exposure for long duration e-learning courses.

ii) Kernel Density Estimation (KDE) analysis



Comparison between KDE plots on time spent on digital tools for learning (before & after Covid) v/s CGPA

In the above graph, we see students had a better chance of getting a higher CGPA if they spent more than 1–3 hours on digital tools for learning after the pandemic. However, in both curves, excessive use of online learning tools leads to a steady decline in academic performance. Thus, intuitively we can say that the excessive use of digital tools may harm the student's academic performance.

iii) Model Selection

a) Decision Tree: It consists of nodes and leaves where data is continuously split in accordance with certain parameters. Another advantage of decision tree is that it requires lesser time for data preparation and is easy to interpret.

Predicting scores through Decision Tree method

	precision	recall	f1-score	support
1	0.41	0.56	0.47	97
2	0.40	0.35	0.38	114
3	0.35	0.43	0.39	93
4	0.53	0.34	0.41	114
accuracy			0.41	418
macro avg	0.42	0.42	0.41	418
weighted avg	0.43	0.41	0.41	418

b) Random Forests: It performs classification by constructing a number of decision trees at training time using bootstrapping, random subsets of features and average voting. It is more robust than decision trees and lesser prone to overfitting.

Predicting scores through Random Forest method

	precision	recall	f1-score	support
1	0.56	0.41	0.47	131
2	0.35	0.40	0.38	99
3	0.43	0.35	0.39	114
4	0.34	0.53	0.41	74
accuracy			0.41	418
macro avg	0.42	0.42	0.41	418
weighted avg	0.44	0.41	0.42	418

c) K-nearest neighbors (KNN): It determines the class of the data point through a majority voting principle. It means that a class label can assigned to a data point based on its distance to its nearest neighbors. It has relatively lesser computation time compared to other classification techniques.

Predicting scores through K-Nearest Neighbors method

	precision	recall	f1-score	support
1	0.75	0.39	0.51	188
2	0.23	0.45	0.30	58
3	0.17	0.33	0.23	49
4	0.48	0.45	0.46	123
accuracy			0.41	418
macro avg	0.41	0.40	0.38	418
weighted avg	0.53	0.41	0.44	418

d) Support Vector Machine (SVM): It is generally used for smaller datasets and hence perform relatively faster. In this technique a hyper plane is drawn which helps to separate two or more different classes. The decision boundary or Hyperplane is estimated by maximizing the distance between different groups. The dimension of the hyperplane depends upon the number of features of the class. Due to its shorter computation time it is often used to predict student performance.

Predicting scores through Support Vector Machine method

	precision	recall	f1-score	support
1	0.56	0.41	0.47	131
2	0.35	0.40	0.38	99
3	0.43	0.35	0.39	114
4	0.34	0.53	0.41	74
accuracy			0.41	418
macro avg	0.42	0.42	0.41	418
weighted avg	0.44	0.41	0.42	418

iv) Cost Estimation using Cocomo Model

The following table describes the different tasks and their LOC values

Sl. No.	Task Name	Duration	Start	Finish	LOC (lines of code)
1	DATA COLLECTION	40 days	8/7/2023	9/29/2023	801
2	SOFTWARE SPECIFICATION	14 days	9/30/2023	10/18/2023	280
3	SOFTWARE DESIGN	34 days	9/30/2023	11/15/2023	716
4	CODING	47 days	11/16/2023	1/19/2024	1600
5	TESTING AND DEBUGGING	24 days	1/20/2024	2/21/2024	780
6	KERNEL DENSITY ESTIMATION	39 days	2/22/2024	4/15/2024	410
7	SCORE PREDICTION	26 days	4/16/2024	5/21/2024	385
				Total LOC	4972
				Effort = a1*(KLOC)^A person months = 2.4*(4.972)^1.05 =	12.92911855
				Development time = b1*(Effort)^B months = 2.5*(12.929)^0.38 =	6.612051041
				Productivity = (KLOC/Effort) = 4.972/12.929 =	0.384558312

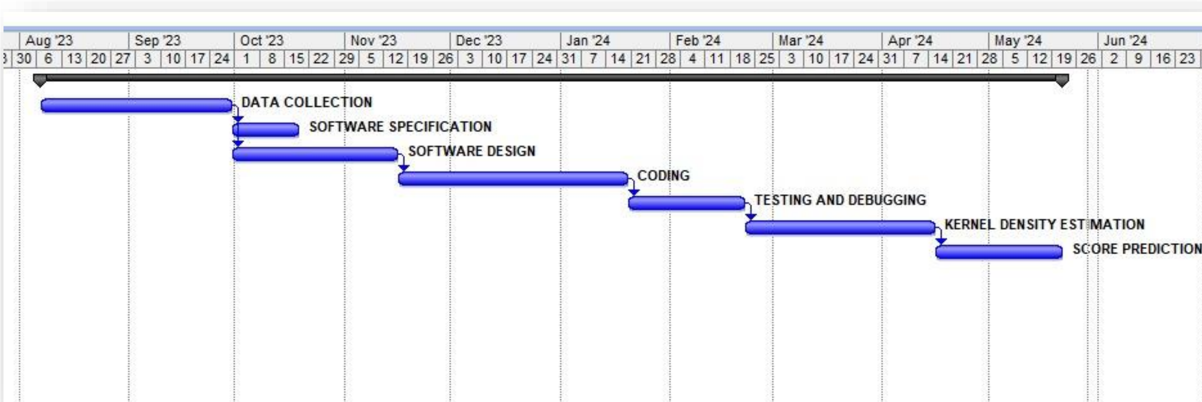
Project	a1	A	b1	B
Organic (Simple)	2.4	1.05	2.5	0.38

From the above calculations we conclude that

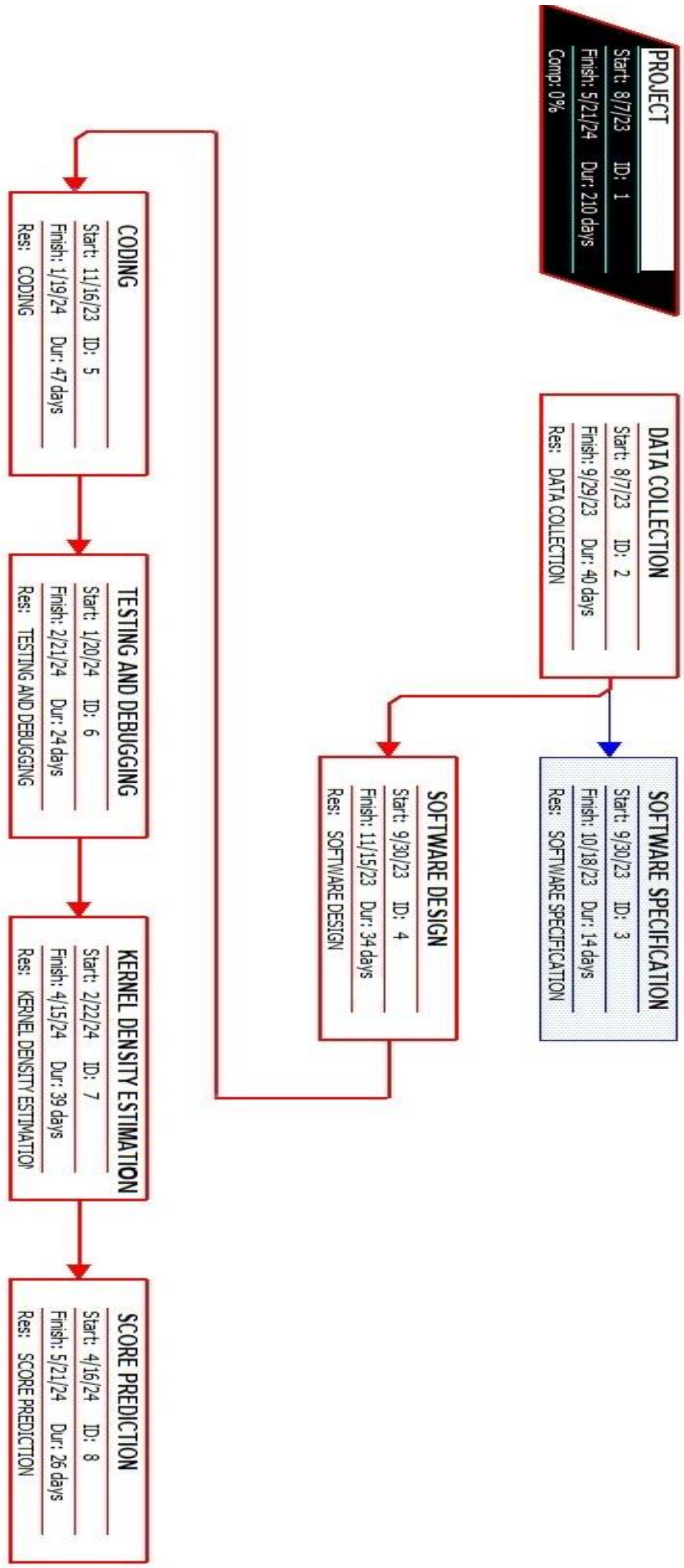
Effort required = 12.929 person-months
Development time = 6.612 months
Productivity rate = 0.3845

a) Gantt chart

The gantt chart of the tasks can be visualized as below:



b) Pert Chart: The pert chart of the tasks can be visualized as follows:



Conclusion and Recommendations

In conclusion, the project "E-learning as a Sustainable Digital Tool" has provided valuable insights into the transformative potential of digital education. Through an in-depth exploration of various facets, including environmental impact, accessibility, and scalability, we have unearthed the ways in which e-learning can contribute to sustainability in the long run.

The environmental advantages of e-learning, such as reduced paper usage and minimized carbon footprint, underscore its role in fostering a more sustainable educational landscape. Moreover, the project has shed light on the inclusivity and accessibility that digital tools bring to education, breaking down geographical and economic barriers.

As we move forward, it is imperative to harness the momentum generated by this project and actively advocate for the integration of e-learning solutions in educational systems worldwide. The collaborative efforts of educators, policymakers, and technology developers are pivotal in shaping a future where e-learning stands as a cornerstone of sustainable education.

In essence, "E-learning as a Sustainable Digital Tool" not only highlights the different impact of digital education on the work environment and accessibility but also serves as a call to action for a collective commitment to building a more sustainable and inclusive educational landscape for generations to come. Through ongoing research, innovation, and collaboration, we have the opportunity to reshape the future of education and pave the way for a truly sustainable and interconnected global learning environment.

Referencing and Appendices

Journal(s) and Book(s):

1. P.S.Bradley, O.L.Mangasarian. "Feature selection via concave minimization and support vector machine." International Conference on Machine Learning (1998): 82-90. 1%
2. Smith, MeCabe and. Unit Operations in Chemical Engg. 4th ed., TMH. 2%
3. Alstete, J.W. and Beutell, N.J. (2004), "Performance indicators in online distance education courses: a study of management education", Quality Assurance in Education, Vol. 12 No. 1, pp. 6-14. 1%
4. Images and other data courtesy jupyter notebook. 2%