**Students’ Adaptability Level in Online Education during COVID-19**

**CISC 525 – Big Data Architectures**

**Final Project Report**

**Instructor:** Sangwhan Cha, PhD

Srikanth Guttikonda

[Introduction 3](#_Toc141984952)

[Motivation 3](#_Toc141984953)

[Data 4](#_Toc141984954)

[Objectives 4](#_Toc141984955)

[System Architecture and Tools 4](#_Toc141984956)

[1. Python 5](#_Toc141984957)

[2. Google Cloud – Hadoop 5](#_Toc141984958)

[3. Spark 5](#_Toc141984959)

[4. Tableau 6](#_Toc141984960)

[Experiments 6](#_Toc141984961)

[1. Python 6](#_Toc141984962)

[2. Google Cloud-Hadoop 6](#_Toc141984963)

[3. Spark 6](#_Toc141984964)

[4. Tableau 7](#_Toc141984965)

[Results 7](#_Toc141984966)

[1. From Python 7](#_Toc141984967)

[2. From Google Cloud-Hadoop 8](#_Toc141984968)

[3. From Spark 8](#_Toc141984969)

[4. From Tableau 9](#_Toc141984970)

[Conclusions 11](#_Toc141984971)

[References 12](#_Toc141984972)

# Introduction

At the beginning of the year 2020, the pandemic of COVID-19 caused a catastrophe on a worldwide scale that had never been seen before, which had a profound effect on the educational system. As a rapid response to the pandemic, educational institutions all around the world rapidly transitioned away from the more conventional in-person instruction and towards education delivered via the Internet. The objectives of this study are to determine the degree to which students were able to adjust to the shift toward online education and to gain an understanding of the difficulties that they encountered during this time. some students are easily adapted to online education but some students are not adapted to online education.

The COVID-19 pandemic forced educational institutions and students to adapt swiftly to online learning. Despite the initial challenges, the majority of students demonstrated remarkable adaptability to the new mode of education.

## Motivation

Challenges Faced by Students:

* Lack of access to appropriate technology and internet connectivity.
* Difficulty in staying motivated and self-disciplined without in-person interactions with peers and teachers.
* Increased distractions at home, leading to reduced focus and productivity.
* Limited social interaction and a sense of isolation negatively impact mental well-being.

The findings highlight the remarkable adaptability of students in response to the sudden shift to online education during the COVID-19 pandemic. While challenges were prevalent, the willingness of students to embrace online learning and the efforts of educators to create conducive learning environments were crucial factors in facilitating adaptability.

* Technological Competence:

The majority of students faced initial difficulties in adapting to online classes due to limited familiarity with digital tools and platforms. However, as time passed, students exhibited remarkable growth in technological competence, with almost all respondents stating they had become more proficient in using digital tools for learning purposes.

* Home Environment:

The suitability of the home environment significantly impacted students’ adaptability to online classes. Students with access to a dedicated study space, stable internet connectivity, and supportive family members demonstrated higher levels of adaptability and performance compared to those with inadequate resources and distractions.

* Self-Discipline and Time Management:

Online learning required students to take greater responsibility for managing their study schedules and adhering to deadlines. While some students thrived in this more independent learning environment, others struggled to maintain discipline, leading to a noticeable divide in academic performance.

* Communication and Interaction:

The reduced face-to-face interaction in online classes led to a decline in social engagement, with some students reporting feelings of isolation and disconnectedness. However, most students adapted by utilizing digital communication platforms to maintain contact with peers and teachers.

## Data

The source of the data is Kaggle. The data set has 14 columns (13 inputs and 1 output) and 1,205 observations. All the columns are categorical, and the target is known (Adaptivity Level). The target variable has 3 classes: Low, Moderate, and High. There are 625 students with Moderate adaptivity, 480 students with Low adaptivity, and 100 with High adaptivity.

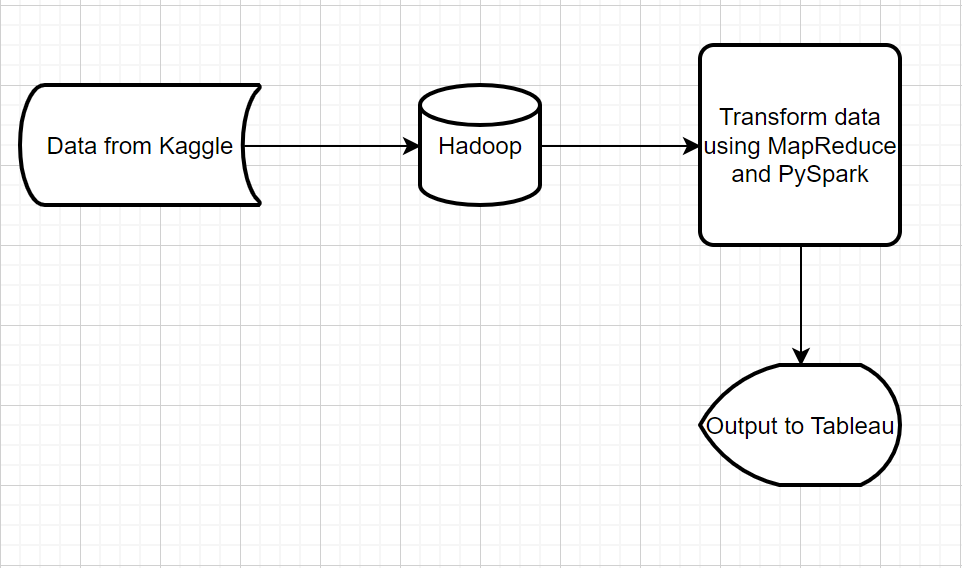
## Objectives

The primary objectives of this project are as follows:

* In this project, Big Data Tools and algorithms will be used to answer the research question and help understand What causes high and low adaptability levels.
* This prediction can be used as input by the management of a school to help determine the class format for every batch of students.
* This project can also help determine if an online school can be an option for some students.

# System Architecture and Tools

Figure 1 is a basic outline of how our Big Data Tools work together and how any data collected flows through the Architecture. It is first collected, in this case from Kaggle, where it is quickly checked using Python for any missing data entries. The data is then uploaded to a Hadoop system, where the data is reduced using MapReduce and Spark technologies into a (Key, Value) pair. It is then transformed into a (Value, Key) pair, and uploaded to Tableau for visualizations.



## Python

Python is a popular and versatile high-level programming language known for its simplicity and readability. It was created by Guido van Rossum and first released in 1991. Python is widely used in various domains, including web development, data analysis, machine learning, artificial intelligence, scripting, and automation. It offers a large standard library and a vibrant community that contributes to numerous third-party libraries and frameworks, making it a favorite choice for developers across the world.

## Google Cloud – Hadoop

Google Cloud is a cloud computing platform offered by Google. It provides a suite of cloud-based services that allow users to build, deploy, and manage applications and services through the Internet. Google Cloud offers a wide range of services, including computing power, storage, machine learning, databases, networking, and analytics, among others. It is one of the major cloud providers, along with Amazon Web Services (AWS) and Microsoft Azure, and it caters to businesses of all sizes.

Apache Hadoop is an open-source distributed computing framework designed to process and store large datasets in a distributed computing environment. Hadoop allows for scalable, reliable, and efficient processing of big data across clusters of commodity hardware. The core components of Hadoop are the Hadoop Distributed File System (HDFS) for storage and the MapReduce programming model for data processing. Hadoop is widely used in big data applications and provides the foundation for various other big data technologies.

## Spark

Apache Spark is another open-source distributed computing framework focused on fast and distributed data processing. It was developed to improve upon some limitations of Hadoop’s MapReduce model, providing faster and more flexible data processing capabilities. Spark allows for in-memory processing, making it much faster than Hadoop for certain workloads. It also offers support for various data processing models, including batch processing, interactive queries, stream processing, and machine learning. Spark has become a popular choice for big data processing and analytics due to its speed, ease of use, and rich feature set.

## Tableau

Tableau is a powerful data visualization and business intelligence software developed by Tableau Software, which is now part of Salesforce. Tableau allows users to connect to various data sources, whether they are spreadsheets, databases, or cloud-based data platforms. Once connected, users can create interactive and visually appealing dashboards, reports, and charts to explore and analyze their data. Tableau’s intuitive drag-and-drop interface makes it easy for both technical and non-technical users to create insightful visualizations and gain actionable insights from data. Tableau is widely used in organizations to help with data analysis, decision-making, and data-driven storytelling.

# Experiments

## Python

Python was used for an initial analysis of the data extracted from Kaggle. Python codes showed the number of students and factors influencing the adaptability level in online education. Also, this tool was used for preprocessing purposes. The presence of missing values was checked during this process.

An initial visualization was created using a pie chart to visualize the different categories and distribution of the target variable (Adaptability Level). This is shown below under the “Results” section.

## Google Cloud-Hadoop

A Hadoop Architecture System was used to help process this data. In this case, we used Google’s “Google Cloud Dataproc”. The two primary reasons for using Google Cloud are because creating a cluster is very easy to do with Dataproc, and when an account is first created, the account gets a $300 credit applied to the account, so we were able to use the Architecture and resources for free. This was also the same tool introduced to us in class, so every group member was familiar with the technology and knew how to run it.

## Spark

Spark was used to reduce the data that we had down to information that was more readable and easier to analyze. We separated all information related to a “High” adaptability rating into one data table, and all information related to a “Low” adaptability rating into a second data table. Specific features, such as a student’s age, location, or internet connection, were then extracted based on these two data tables for analysis. The MapReduce functions that are built into PySpark were then used to get this information into (Key, Value) pairs.

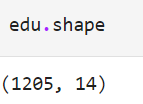
## Tableau

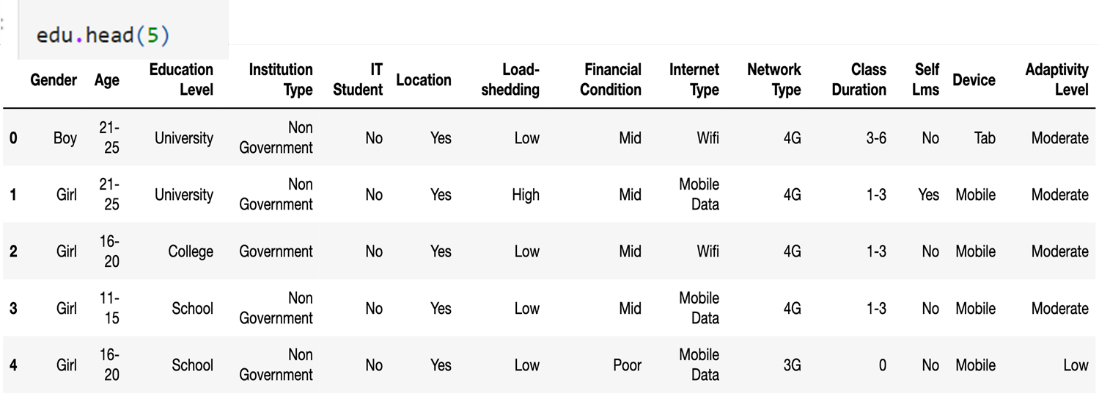
Tableau was used to visualize outputs obtained from Spark regarding low and high adaptability levels.

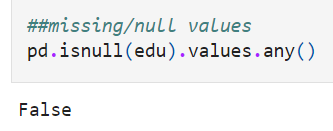
# Results

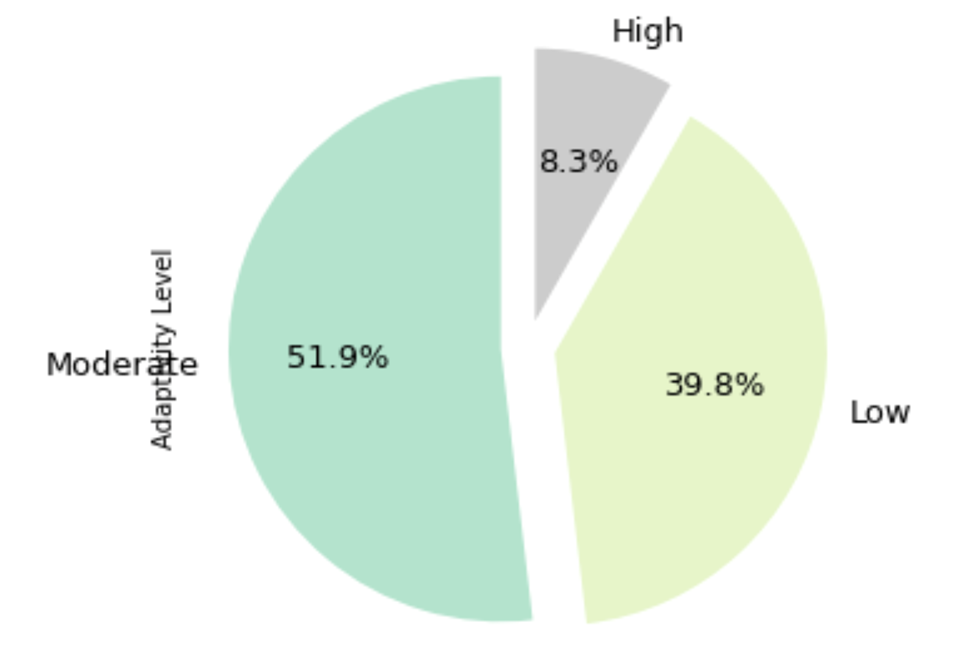
## From Python

The data has 1,205 students and 14 factors. The dataset does not include missing or duplicate values. The pie chart created shows that 8.3% of students adapted to online education highly and 39.8% had the hardest time adapting to online education. The data has 1,205 students and 14 factors. The dataset does not contain missing values. The pie chart created shows that 8.3% of students adapted to online education highly and 39.8% had the hardest time adapting to online education.









## From Google Cloud-Hadoop

The .csv file that is collected from Kaggle is uploaded to our Hadoop System. Since we used Google’s Dataproc technology, uploading the file to the system itself was as simple as clicking two or three buttons, and the file showed up in our system, as shown below as the “students.csv” file.

A black background with white text

Description automatically generated

This file was then uploaded to the HDFS system, where PySpark could be used. Also, this Hadoop system is comprised of one Master Node and 4 Worker Nodes.

## From Spark

When the CSV file was loaded into PySpark, it was loaded in as a DataFrame Object, therefore most of the processes done on the data are DataFrame functions. The code snippet below shows what code was run and how. First, the data is filtered out based on whether it’s a “High” or “Low” adaptability rating, and then a feature is chosen. The data that is found is then transformed into a (Key, Value) pair based on the number of times the specific type of feature appears. A great example is the second screenshot below, which shows the number of times each of the three types of “devices” show up. In other words, it shows how many students of each of the corresponding adaptability levels had each type of device as their primary device for school.

A screen shot of a computer program

Description automatically generated

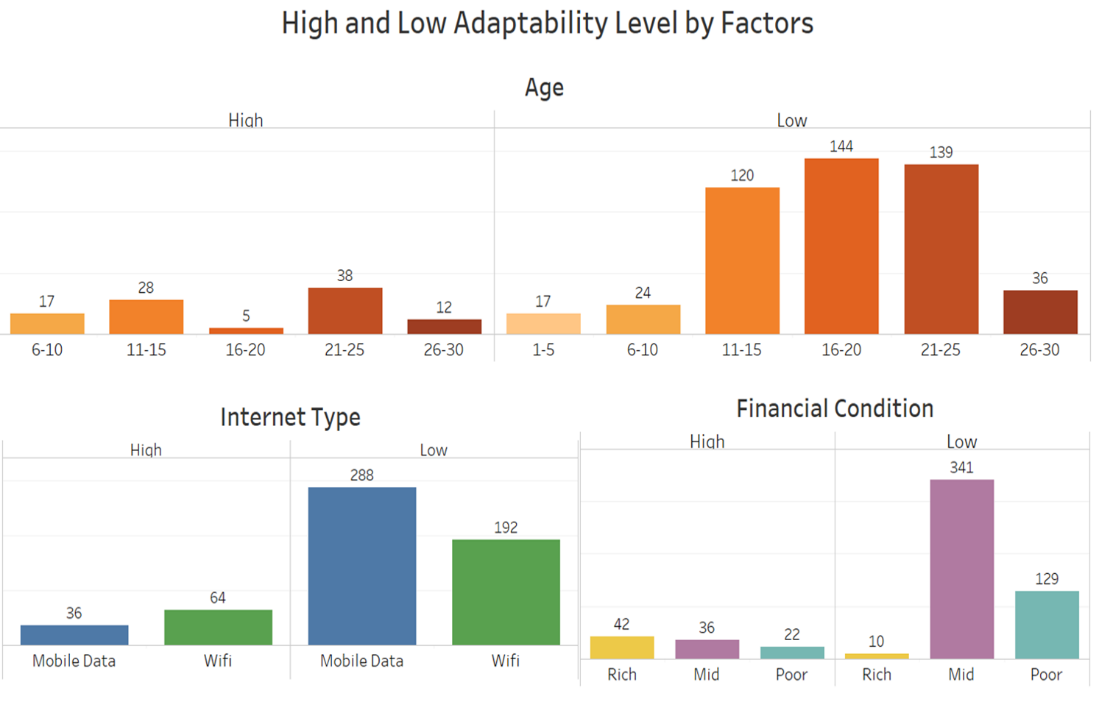
A computer code on a black background

Description automatically generated

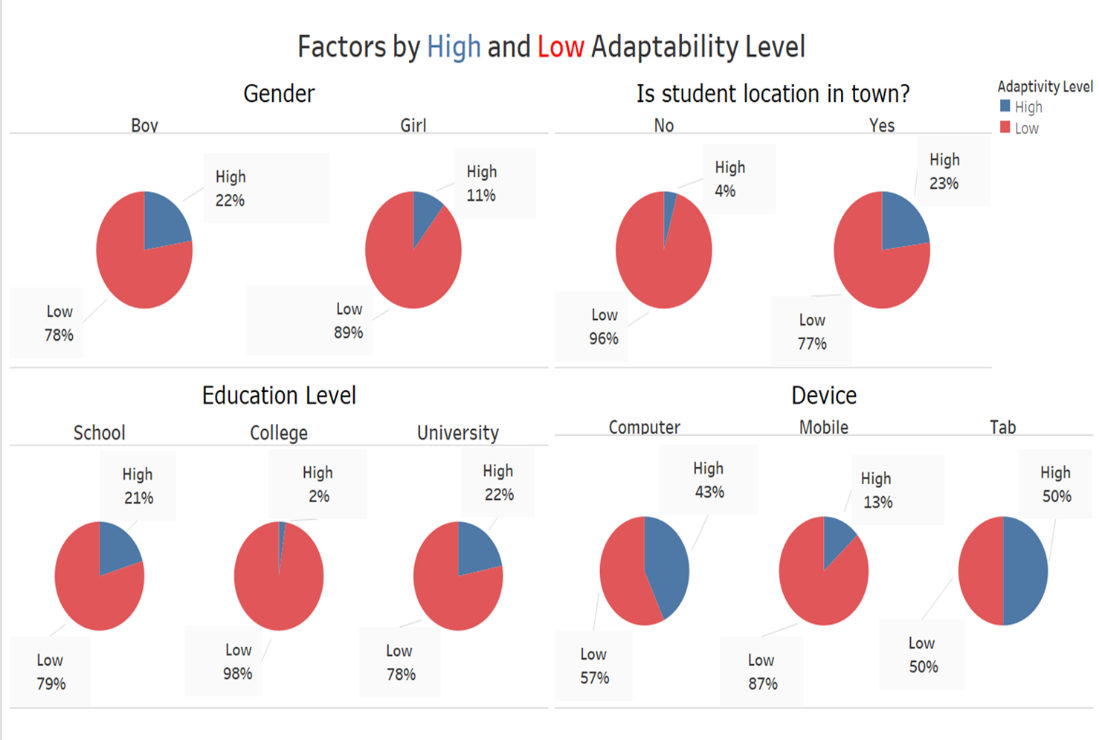
Any feature that is run will have a similar-looking output. This process can also be automated by allowing the “feature” variable to equal a list or array of features you would want to analyze, and running some form of a loop on it. These (Key, Value), or in this case, (Value, Key), pairs are then fed into Tableau to produce visualizations of the data.

## From Tableau

Two dashboards in Tableau were created. The first dashboard was created by showing 3 factors: Age, Internet Type, and Financial Condition. Different colors and bar charts were used to represent categories. The Age graph shows that 38 students from the 21-25-year-old group considered that their adaptability level to online education was high. On the other hand, 11-15, 16-20, and 21-25-year-old groups had the hardest time adapting. The Internet Type graph shows that the students with low adaptability levels were using Mobile Data. The Financial Condition graph shows that Mid was the category with the lowest adaptability level.



The second dashboard was created using a different visualization technique. Instead of using different colors to distinguish between categories, we use different colors to represent the Adaptability Level categories of interest in our project (High and Low). The visualization shows that students living in town had a better chance of adapting than those who live in rural areas. The Education Level graph shows that students enrolled in a School or University have higher percentages of high adaptability compared to students in College. The Device graph shows that students using Mobile devices had the hardest time adapting to online education.



# Conclusions

The COVID-19 pandemic presented a unique opportunity to assess student adaptability to online education. While challenges were evident, students showed resilience and developed coping strategies to overcome them. Institutions and educators should use these insights to improve the online learning experience by providing better support, enhancing communication, and incorporating diverse learning resources. By doing so, the education system can be better prepared to face similar challenges in the future while fostering a conducive learning environment for all students.

* Students using Mobile devices and Mobile Data tend to have a lower Adaptability level.
* Students living in the town seem to have a better chance of adapting to Online learning than those who live in rural areas (internet connection is Limited).
* 11-15- and 16–20-year-old groups had the hardest time adapting.

# References

Suzan, M. H., Samrin, N. A., Biswas, A. A., & Pramanik, A. (2021, July). Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE