****

**PERFORMING A SENTIMENT ANALYSIS TO CHECK THE EFFECTIVENESS OF THE REMOTE TEACHING AND LEARNING DURING THE COVID-19 PANDEMIC USING NAÏVE BAYES.**

BY

MAKUWA THEMBI REBOTILE

201730087

A RESEARCH PROPOSAL SUBMITTED FOR THE DEGREE OF HONOURS SCIENCE (COMPUTER SCI) IN THE DEPARTMENT OF COMPUTER SCIENCE, SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCE, FACULTY OF SCIENCE AND AGRICULTURE, UNIVERSITY OF LIMPOPO, SOUTH AFRICA.

SUPERVISOR: MS OLEBOGENG NOJILA

CO-SUPERVISOR:

OCTOBER, 2022

# Dedication

I dedicate this dissertation to my parents (**Ntjemeng and Moureen)**, my 2 brothers and 4 sisters **(Thabo, Nelson, Meisie, Ronnica, Khomotso and Nelly**), all my classmates and most important to myself.

# Declaration

I certify that all work done for this dissertation was done by me and this project does not include any material that was written or published by a third party unless it was properly cited with thorough and accurate reference. Every significant aid source has been acknowledged.

12 October 2022

**……………………. …………………….**

**Signature Date**

# Acknowledgements

Prior to all else I want to express my sincere appreciation to the Almighty God for the guidance throughout everything, my supervisor (Ms OH Nojila) for her encouragement, advice, and unwavering patience during this assignment, and my sister (Khomotso) for always being my pillar of strength. It is also noted that without the assistance of the following individuals, this project would not have been successful:

1. The University of Limpopo, Faculty of Science and Agriculture, School of Mathematical and Computer Science under the department of Computer Science for instituting me.
2. The Faculty dean, school director, the Computer science HOD and as well as all the computer science lectures for the 2022 academic year for their leadership, support and undying patience.
3. My parent for financial support.
4. Mr PT Malatjie for being the core coordinator of the research.

# Abbreviations

NB- Naïve Bayes

ML- Machine Learning

WHO- World Health Organization

API- Application Programming Interfaces

NTLK- Natural Language Toolkit

Res- Regular Expressions

SDIT-State Department of Information Technology

URLs-Uniform Resource Locators

BBA-Bachelor of Business Administration

BCA-Bachelor of Computer Applications

BSC-Bachelor of Science

BA-Bachelor of Arts

ERL­- Emergency Remote Learning

# List of Figures

[**Figure 1:** Details of how to create a Twitter API account 28](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851261)

[**Figure 2:** Details of Twitter API keys 29](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851262)

[**Figure 3:** Details of importing Libraries. 30](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851263)

[**Figure 4:** Details how to make a seaech term on Twitter](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851264) 30

[**Figure 5:** Initial Dataset 31](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 6:** Initial dataset in a DataFrame 32](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 7:** Details Cleaning and Preprocessing 33](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 8:** Polarity nd Subjectivity 35](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 9:** WordCloud visualization 36](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 10:** Sentiment Analysis Model 37](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 11:** Distribution Plot 38](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 12:** Bar graph Results 39](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 13:** Pie chart Results 40](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

[**Figure 14:** Details probability Satisfaction Results 41](file:///C:\Users\201730087\Downloads\Hlaraka%20Thapelo.docx#_Toc89851265)

# List of Tables

Table 1: Literature Review Matrix

Contents

[Dedication i](#_Toc117553750)

[Declaration ii](#_Toc117553751)

[**Acknowledgements** iii](#_Toc117553752)

[Abbreviations iv](#_Toc117553753)

[List of Figures v](#_Toc117553754)

[List of Tables vi](#_Toc117553755)

[Abstract ix](#_Toc117553756)

[Chapter 1 Proposal 1](#_Toc117553757)

[1.1. Introduction 1](#_Toc117553758)

[1.2. Problem Statement 2](#_Toc117553759)

[1.3. Rationale or Motivation 3](#_Toc117553760)

[1.3.1. Aim 3](#_Toc117553761)

[1.3.2. Objectives 3](#_Toc117553762)

[1.4. Literature Review 3](#_Toc117553763)

[1.5. Methodology and Analytical Procedure 4](#_Toc117553764)

[1.5.1. Data Collection 4](#_Toc117553765)

[1.5.2. Data Cleaning 4](#_Toc117553766)

[1.5.3. Tools and Experimental Setup 5](#_Toc117553767)

[1.5.4. Data Analysis 5](#_Toc117553768)

[1.6. Scientific Contribution 6](#_Toc117553769)

[1.7. Chapter Summary 6](#_Toc117553770)

[Chapter 2 Literature Review 7](#_Toc117553771)

[2.1. Introduction 7](#_Toc117553772)

[2.2. Literature Review Matrix 7](#_Toc117553773)

[2.2.1. Public opinions about online learning during COVID-19 14](#_Toc117553774)

[2.2.2. Challenges in remote teaching and learning 16](#_Toc117553775)

[2.2.3. Benefit of online teaching and learning 20](#_Toc117553776)

[2.2.4. Ways to improve online teaching and learning 23](#_Toc117553777)

[2.3. Chapter Summary 24](#_Toc117553778)

[CHAPTER 3 Research Methodology 25](#_Toc117553779)

[3.1. Introduction 25](#_Toc117553780)

[3.2. Data collection 25](#_Toc117553781)

[3.3. Data Cleaning 25](#_Toc117553782)

[3.4. Model training and testing 26](#_Toc117553783)

[3.5. Data Visualization 27](#_Toc117553784)

[3.6. Techniques to evaluate 28](#_Toc117553785)

[3.6. 1.Twitter API Application 28](#_Toc117553786)

[3.6.2. Twitter API keys 28](#_Toc117553787)

[3.6.3. Importing Libraries 29](#_Toc117553788)

[3.6.4. Search a Term 29](#_Toc117553789)

[3.6.5. Initial Dataset 29](#_Toc117553790)

[3.6.7. Data Frame 30](#_Toc117553791)

[3.6.8. Data Cleaning and Preprocessing 31](#_Toc117553792)

[3.7.1. Extract features and train a classifier 31](#_Toc117553793)

[3.8. Chapter Summary 32](#_Toc117553794)

[Chapter 04 Results and Analysis 33](#_Toc117553795)

[4.1. Introduction 33](#_Toc117553796)

[4.2. Polarity and Subjectivity 33](#_Toc117553797)

[4.3. WordCloud visualisation 33](#_Toc117553798)

[4.4. Sentiment Analysis model 34](#_Toc117553799)

[4.5. Distribution Plot 34](#_Toc117553800)

[4.6. Bar Graph results 35](#_Toc117553801)

[4.7. Pie chart 36](#_Toc117553802)

[4.8. Objective 2 36](#_Toc117553803)

[4.8.1. Factors that affect the effectiveness of online teaching/learning 36](#_Toc117553804)

[4.9. Objective 3 37](#_Toc117553805)

[4.10. Chapter Summary 37](#_Toc117553806)

[Chapter 05 38](#_Toc117553807)

[5.1. Introduction 38](#_Toc117553808)

[5.2. Answers per objective 38](#_Toc117553809)

[5.2.1. Objective 1 38](#_Toc117553810)

[5.2.2. Objective 2 38](#_Toc117553811)

[5.2.3. Objective 3 38](#_Toc117553812)

[5.3. Future Work 39](#_Toc117553813)

[5.4. Limitations 39](#_Toc117553814)

[5.5. Chapter Summary 39](#_Toc117553815)

[References 40](#_Toc117553816)

[Appendix 42](#_Toc117553817)

# Abstract

Despite the widespread use of emergency remote learning (ERL) and the community being divided on whether or not technology is ready during the COVID-19 pandemic in higher education, there is enough knowledge of educational technology for this innovative environment in crisis. The goal of this study was to perform a sentiment Analysis, using Twitter opinions to test the effectiveness of remote teaching and learning in the course of COVID-19 pandemic. While some research indicated that students' and lecturers' excitement is the key to effective online learning, others discovered that a lack of resources has a negative impact on both student and lecturer performance. The study that was done was based on information extracting text mining and feelings on Twitter data analysed utilizing Naïve Bayes procedure. 41% of respondents in that analysis were positive, 23.2% were negative, and 35.8% were neutral. Several tweets revealed that "innovation" and "covid" were the most often used words.

# Chapter 1 Proposal

## Introduction

The COVID-19 pandemic hit the world harshly and unexpectedly after it was declared a universal pandemic on March 11, 2020 by the World Health Organisation (WHO) (Ghebreyesus, 2020). As a result, the educational sector has changed, and the way in which students are instructed and connect with instructors has changed so much with the internet, live-stream classes, cellphones, and websites. Numerous universities relied on the manual method of teaching and learning, which cannot be used under current COVID-19 regulations, especially in most qualifications (Reimers, 2020). In this case, universities, therefore, had to look for an emergency alternative method to get the teaching and learning to continue and to save the academic year. One of such alternative is remote teaching/learning procedure. The remote teaching and learning method is the process of learning and teaching at a distance, using either a smartphone or a laptop and access to the internet connection (Anon., 2020). This method was previously used by institutions that often offer distance learning. Although students are acknowledgments e-learning materials, most of them are still disadvantaged due to the connectivity issues faced, and the maintenance required by these learning gadgets.

Education quality has been improving in South Africa, off a lower base, according to international testing programs (Stellenbosch Economic Working Papers, December 2020). Along with knowledge of appropriate pedagogical methods for the online environment, creating high-quality live-stream courses involves expertise, resources, and other factors (Kaushal Kumar Bhagat, 2021). But these resources have been in short supply during the pandemic. As vital as it is to try to predict the pandemic's effects, the current high levels of uncertainty mean that any estimates must be constantly questioned and re-calculated (Wu, 2020). Under-privileged students are compromised when they run out of data to continue the learning, and can’t access the campus, teachers are also compromised because they must now work from home which becomes a challenge most of the time.

This study will focus on testing how effective online teaching and learning is during COVID-19, and whether this method should continue being carried out or not, based on the Twitter opinions.

## 1.2. Problem Statement

The remote teaching and learning method was implemented to minimize the effects of COVID-19. However, the utilization of educational technology remains difficult to attain. The first problem is understanding the subject materials, not all students may be able to understand the online reading material’s content. The ability of the instructor to employ technology in online learning, not all teachers are computer or gadget savvy enough to participate in online learning activities (Arambepola, 2020). Even though student’s understanding is tested through assignments, quizzes and tests, but there is no much effort, enthusiasm put in preparing, because there is an advantage of searching for answers from external sources (Özüdoğru\*, 2021). When it comes to competence (both cognitive and non-cognitive skills) and availability to assist their children with homework at home during the lockdown, parents from varied socioeconomic backgrounds may differ (Di Pietro, 2020). South Africa often faces load-shedding, which is a negative contributing factor to online teaching and learning. The current remote teaching and learning method requires higher internet penetration, and since it is conducted at a distance teachers were taken off guard and unprepared (Özüdoğru\*, 2021). In addition to knowledge of the appropriate pedagogical approaches for the online environment, creating high-quality online courses demands resources and talents.

According to Davis, Gough, and Taylor (2019), based on literature " instructors obstruction to online learning may be related to expectations identification, feedback delivery, and human relationships, whereas student obstacles may be related to expectations misinterpretation, time management, as well as interpersonal interactions in their study. O'Doherty, Dromey, Lougheed, Hannigan, Last, and McGrath (2018) claimed that obstacles to online learning in medical training could include the absence of institutional plans and support, time restrictions, a lack of technical skills, poor infrastructure, and the attitudes of all parties involved. (Gul Ozudogru, 2021).

## 1.3. Rationale or Motivation

The goal has always been one, to deliver quality education and produce graduates who have necessary skills that can be applied in the corporate World. There are some factors that motivate this study, one which is, there is still uncertainty as to whether the method should continue or not. Network plays a major role in remote teaching and learning, but even as of today there are still students together with instructors who reside in areas that do not have fast internet access to enable their participation in remote teaching and learning. Statistics show that South Africa falls within the top 5 of countries that have high crime rate globally (Anon., 2021), therefore even though students and teachers are given materials that will enhance their participation in the method, they are still at risk of losing those materials.

### 1.3.1. Aim

The Aim of this study is to perform a sentiment Analysis, using tweeter opinions to test the effectiveness of remote teaching and learning during COVID-19 pandemic.

### 1.3.2. Objectives

Listed below are the goals of this investigation:

* To classify the user’s opinion about distance learning as positive, negative, and neutral during the COVID-19 pandemic in South Africa.
* To identify factors that affect the effectiveness of online learning.
* To determine the probability of each factor on a student’s performance, using Machine Learning.

## 1.4. Literature Review

Kaushal Kumar Bhagat**,** et. al (2021) conducted a study to perform a sentiment analysis to investigate public opinions about online learning during COVID-19. In the initial stages of the pandemic, they examined the web news and blogs. Related work, such as web scraping was used to extrapolate a lot of information from websites to accommodate various scenarios. The potential advantages of online learning were studied. All these variables were found to be motivating teachers and students to include online learning into their academic programs.

In their research, they used lexicon-based techniques and the dictionary-based methodology to perform the analysis of the articles that were scraped from the web using sentiment.

After all, there are a lot of factors that can influence online learning, and affect teachers and learners both in negative and positive ways. These external factors, jointly with where an individual resides, should significantly affect online learning and enhance the effectiveness of remote learning.

However, this study focused more on positive factors and revealed online in public digital media are perceived more favorably yet cautiously, with a low polarity value. Even though the study labels other factors that can affect online learning it does not elaborate on how it will affect the current remote teaching and learning.

## 1.5. Methodology and Analytical Procedure

The first step will involve collecting data. The next step will be data pre-processing, followed by performing a sentiment Analysis on the collected data. Then finally perform a statistical data analysis and conclude. Machine learning algorithms, programming language, together with other tools will be deployed to complete this study.

### 1.5.1. Data Collection

In this study, data will be collected from Twitter as a data source. Because tweets reflect their opinions about their experiences using products, services, and new trends, Twitter users are helpful for current, real-world events like the COVID-19 epidemic (Arambepola, 2020).

### 1.5.2. Data Cleaning

Because there are irrelevant elements in tweets, as well as various language patterns, it is necessary to pre-process the data in order to standardise specific tokens of tweets. Data pre-processing will be performed to transform raw dataset into an understandable format and to employ techniques like the following you may eliminate unnecessary noise from the textual data.

* Convert all texts into lower case
* Eliminate all URLS
* Remove all the hashtags, the user mentions, and emoji
* Remove all the punctuations.
* Remove all the numbers.
* Remove parentheses.

### 1.5.3. Tools and Experimental Setup

Naïve Bayes will be used as the machine learning algorithm to classify data into either positive, negative, or neutral. This method is technically reasonable in relation to consistent categorisation of measurements and data. The Naïve Bayes classifier uses the principle of conditional probability, as given by the Bayes Theorem (Simplilearn, February 2022). For this Sentiment Analysis, we will use Naive Bayes (NB) to compute posterior probabilities (M.S. Farooq, 2011)

The programming language Python will be deployed for text pre-processing. And some python libraries.

Any computer with core i5, 64GB ram computer system, and deployed the Naïve Bayes algorithm

### 1.5.4. Data Analysis

Naïve Bayes will use the method of conditional probability to classify data, and validate the results. We will be compiling a dictionary of these words and collecting the frequency of each word in good, negative, and neutral tweets in order to obtain the probability statistics for the words (Khandewal, 2020). To determine a tweet's emotional content, multiply the likelihood ratio of each word by the number of words in the tweet.

For this study we will use word polarity and subjectivity to determine the sentiment Analysis. To determine the polarity of a word we use a lexicon, such as a dictionary (Marr, 2018).

Python libraries such as the matplotlib and seaborn will be used to visualize the results (Chu, 2021).

## 1.6. Scientific Contribution

The findings of this study will offer a conclusion on the efficiency of distance learning and teaching and whether this method should continue being carried out or not, based on the results found.

## 1.7. Chapter Summary

In this chapter the proposal of the whole study has been summarised, this chapter clearly outlines the background of the study, the motivation behind it, the problem statement, the aim of the study, and how this study will be carried out.

# Chapter 2 Literature Review

## 2.1. Introduction

This section seeks to review the effectiveness of the currently existing online learning method. According to Silvana di Gregorio (2020). No matter the subject or discipline, all dissertations must include literature reviews.

The findings of the research studies focused that students and lecturers performance in online teaching and learning is affected by lack of resources, because this was an urgent and immediate method they had to adopt to. But this should not be considered as a constraint because there are other factors like internet availability, perceived value, course material, course layout, and usability that can affect the performance of students and lecturers (Thi Tinh Thuong Pham, 2021)

## 2.2. Literature Review Matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Topic | Tools and Technologies | Approach | Reference | Relevance |
| perform a sentiment analysis to investigate public opinions about online learning during COVID-19 | web scraping to extract articles. | Sentiment analysis is carried out using lexicon-based procedures using a dictionary-based methodology. | Kaushal Kumar Bhagat**,** Sanjaya et. al. (2021) | This study does not elaborate on how factors affect the current remote teaching and learning. |
| Analysing the tweets about distance learning during COVID-19 Pandemic using Sentiment Analysis | Twitter dataset | Performed experiment analysis, data pre-processing, classify tweets and performed a statistical analysis. | Nimasha Arambepola (2020) | This study only focused on 6 countries Worldwide. |
| Performing a sentiment Analysis during COVID-19 Indonesia Pandemic based on twitter data | Naïve Bayes ML algorithm | Twitter sentiment analysis and document-based text mining with the use of NB algorithm. | Syafrida Hafni Sahir, et. al. (2021) | This study shows that with the use of different algorithms will probably result in better and more consistent performance. |
| Performing a sentiment analysis on tweets about online learning during COVID-19 in Saudi Arabia | Arabic Twitter dataset and Python libraries. | Measured sentiments characteristics using TextBlob | Asma Althagafi, et. al. (2021) | This study did not have sufficient resources to analyse  Arabic language. |
| Investigating the challenges faced during online teaching and learning in one University in Lesotho during the COVID-19 crisis | Telephone to conduct interview for data collection | Quantitative approach, conducted interview | Julia Chere-Masopha (2021) | This study focused on a small sample size, which gave inaccurate results. |
| Predicting student’s satisfaction of emergency remote learning in higher education during COVID-19 using Naive Bayes | 1410 students from different faculties, Naïve Bayes | Collected Twitter dataset from https://TrackMyHashtag.com | Indy Man Kit Ho, et.al (2021) | This study sfocused on one machine learning algorithm. |
| Analysing the student’s difficulties during online learning in speaking for professional context during COVID-19 | They used WhatsApp to carry out their study. | Descriptive qualitative technique approach such as questionnaire and interview | Sri Ariani, Tawali (2021) | This study does not elaborate how the problems stated affect the current online teaching and learning and therefore urges for more research on student’s methods for dealing with these problems and adapting to online learning |
| Exploring challenges faced by lecturers in conducting online teaching during COVID-19 | 30 lecturers | Thematic analysis to carry out the survey of questions | B. Ramayah, R. Kuma (2020) | This study focused more on undergraduate programs. |
| Investigating the effectiveness of online learning during COVID-19. | Google forms to conduct a questionnaire | Questionnaire approach. | Vaibhav Gallani (2020 | This study used many undergraduate students as compared to postgraduate students |
| Determining the effectiveness of online learning during the COVID-19 pandemic | whatsapp, Webex, and Google classroom  Sample of 39 students for quantitative, and 3 students for qualitative | Both quantitative and qualitative approach | Suprianto, et.al (2020) | This study has few factors that influence the effectiveness of online learning, and does not detail how these factors affect the current online learning approach. |
| Determining the effectiveness of online learning during the COVID-19 pandemic | Internet to conduct a questionnaire. | This study uses a quantitative approach | Awal Bahasoan, et.al (2020) | This study did not have sufficient resources to carry out this research. |
| Describe the effectiveness of online learning using Undiksha e-learning in the physical education and health study program during the COVID-19 pandemic | 120 students from the contact education method program, a questionnaire to collect data | This study uses a quantitative approach such as a questionnaire. | I Made Satyawan, et.al (2021) |  |
| Understanding and explore the strategies used by teachers of the SDIT Rabbi Radhiyya Curup to increase student’s interest in learning and to express student’s responses in online learning during the COVID-19 pandemic | Principal, vice-principal, teachers and students were used to conduct an interview | This study used a qualitative approach  An interview was conducted | Sutarto Sutarto, et.al (2020) |  |
| Evaluating the strategies to improve teacher ability in using the madrasah e-learning application during the COVID-19 pandemic | 36 teachers who teach Mardrash Aliyah Plus Nurul Islam | 3 stages were followed.  One being a preparations stage, 2nd implementation stage and lastly evaluation stage. | Deli Reyan Rizaldi (2021) | The results of this study showed that teachers need the mentoring activities in order to understand the use of ministry of religion online learning applications with different supporting features |

### 2.2.1. Public opinions about online learning during COVID-19

* Kaushal Kumar Bhagat**,** Sanjaya et. al. (2021) conducted a study to perform a sentiment analysis to investigate public opinions about online learning during COVID-19. They analysed blogs and online news during the early days of the pandemic. Related work, such as web scraping was used to extract enormous volumes of information from websites to different scenarios. The potential advantages of online learning were studied. All these variables were found to be motivating teachers and students to include online learning into their academic programs.

In their research they employed the lexicon-based techniques' dictionary-based approach to perform the analysis of the articles that were scraped from the web using sentiment.

Google and DuckDuckGo were used to retrieve 154 articles about online learning from news and blogging websites (Kaushal Kumar Bhagat, 2021). In general, they discovered that blogs were more positive than newspaper pieces, with over 90% of the articles being favorable and the remaining few being somewhat negative.

Although, this study focused much on positive factors and revealed more online in public digital media are seen favorably yet cautiously, having a low polarity value. Even though this study labels other factors that can affect online learning it does not elaborate on how it will affect the current remote teaching and learning.

* Nimasha Arambepola (2020) conducted a study to analyse the tweets about distance learning during COVID-19 Pandemic using Sentiment Analysis. The effectiveness of remote teaching and learning was thought to be impacted by a number of variables, including the accessibility of internet resources and the user's financial situation. For this study an experimental analysis was performed, with the use of twitter data-set which contain 202645 tweets. The analysis of tweets consisted of 3 stages, 1st which was to pre-process the data with the aim of removing unrelated tweets, 2nd ways to classify tweets into either positive, negative or neutral based on their polarities and the 3rd step was to perform a statistical analysis to find the connection between user experiences with distant learning in various nations.

These results found showed that 54% sentiments were positive, while 30% were negative and the remaining 16% were neutral. The study concluded that with sufficient internet and the availability of other facilities, distance learning is the best suitable approach in continuing the system of education.

Even though this study focused on both developed and underdeveloped countries, where developed countries were found to have high internet usage which resulted in positive opinions about distance learning, it has focused only on 6 countries World-wide. This means that the obtained results will mainly focus on the selected countries, and if all the countries were to be considered the results may lead to different changes.

* Syafrida Hafni Sahir, et. al. (2021) conducted a study to perform a sentiment Analysis during COVID-19 Indonesia Pandemic based on twitter data. They discovered that technology is the primary element influencing how successful distance learning and teaching is. In their research they used sentiment analysis on data from Twitter and document-based text mining that they analysed with the use of Naïve Bayes method.

The results found showed that 25% sentiments were positive, 74% sentiments were negative and 1% sentiment was neutral. Despite the fact that there are numerous elements that can influence how effective online teaching and learning is, this study only focused on one.

And therefore this study shows that with the utilisation of various algorithms it is more likely to obtain much reliable performance.

* Asma Althagafi, et. al. (2021) conducted a study to perform a sentiment analysis on tweets about online learning during COVID-19 in Saudi Arabia. They used the emotional reactions of the public on online learning as a contributing factor for the efficiency of online learning. Their paper studied the attitude of people of Saudi Arabia towards online learning.

They collected Arabic tweets that were posted in 2020 to perform a sentiment analysis. 10455 Arabic tweets were collected for sentiment analysis, sentimental characteristics such as polarity and subjectivity were measured, using TextBlob. In their results it has shown that most people have reacted to online learning in a neutral manner.

A sentiment analysis determines whether a given sentence contains either positive, negative or neutral attitudes (Arambepola, 2020). Therefore equal sentences that consist of these 3 emotions must be taken into consideration to get fair results.

In their study however most of the tweets collected did not express negative nor positive emotions. This means that the obtained results might change if other emotions were equally considered. Another challenge was that there was a lack of resources to be able to analyse the Arabic language. This research could yield much better results with future work in understanding people’s emotions towards different online education platforms.

### 2.2.2. Challenges in remote teaching and learning

* Julia Chere-Masopha (2021), conducted a study to investigate the challenges faced during online teaching and learning in one University in Lesotho during the COVID-19 crisis. This study was focused specifically on the kind of challenges experienced by students and how students tackled these challenges.

This study used the qualitative method approach where telephone semi-structured interviews were used as a tool to collect data from twelve undergraduate students from the Education faculty.

According to this study, challenges faced by students were persuaded by their personal traits, educational difficulties and how the University responded. This study concluded that the way in which online teaching and learning was carried out was not favourable to effective online lecturers and students participation. Lecturers were caught unprepared and also students were not receiving quality learning, while the existing digital conditions excluded some students from learning. This study further suggests that the University should invest more in digital infrastructure, processes and as well as strategies that would enhance students and lecturer’s online teaching and learning experience.

There are many other factors that can act as a challenge in one’s online teaching and learning prior to those mentioned in this study. It is known as a fact that students come from different backgrounds, meaning even their personal attributes are different, therefore some students may have positive attributes towards this method. This study only focused on 12 students in one faculty, statistically 12 is referred to as a small sample size which gives inaccurate results. Therefore the findings of this research will not really help one understand the challenging factors faced by students during online learning.

* Indy Man Kit Ho, et.al (2021), conducted a  conducted a survey study to Predicting student’s satisfaction of emergency remote learning in higher education during COVID-19 using Naive Bayes.

In their research they conducted a survey that included 1410 students from science, technology and as well as engineering programmes, this was done through both private and public Universities in Malaysia.

They found that there were a lot of difficulties faced by both students and lecturers throughout the online teaching and learning, this includes difficulties in access to decent gadgets and strong internet penetration to their study environments. It was reported that more than 80 percent of students think online learning makes technical courses such as computer programming, and material science more difficult. They further suggested Universities must create ways to provide online education with a local component to present and potential students in order to stay relevant.

  However this study focused on one machine learning algorithm.

* Sri Ariani, Tawali (2021) conducted a study to analyse the student’s difficulties during online learning in speaking for professional context during COVID-19. Factors such as whatsapp were used to conduct the online classes.

A descriptive qualitative technique approach was used to perform this study. They conducted an online questionnaire as well as an interview as the method of collecting data from 45 students in the 2nd semester of the Department of English at the Universitas Pendidikan Mandalika in Mataram, West Nusa Tenggara,and Indonesia who participated in a speaking for professional context lesson. The questionnaire had twenty questions. Students were asked to reflect on their learning challenges during the pandemic, and it was assumed that they would give more honest answers. The Strongly Agree, Disagree, Agree and Strongly Disagree scales were used as options in the questionnaire.

The findings of this research showed different difficulties faced by students in online learning. In the Speaking for Professional lesson, there were a total of 5 problems that surfaced throughout online learning. The 1st problem had to do with using YouTube videos as a learning resource. The 2nd problem was how to deal with the students' inability to elaborate on the lecturer's topics in an online class. The 3rd problem was students' approach in completing given projects by lecturers. The 4th problem was dealing with student’s inability to access ZOOM Meetings. The last problem was the learning tools and supporting facilities, such as digital devices and internet connection in online environments.

However this study does not elaborate how these problems affect the current online teaching and learning and therefore urges for more research on student’s methods for dealing with these problems and adapting to online learning.

* B. Ramayah, R. Kuma (2020) conducted a study to explore challenges faced by lecturers in conducting online teaching during COVID-19. This study was conducted in 2 Universities in Malaysia. Various challenging factors towards teaching throughout the COVID-19 pandemic in Malaysia were studied, also the results and the analysis of this study detailed the lecturer’s perception towards online teaching.

This study used both the qualitative and quantitative methods and used thematic analysis to analyze the survey questions. A survey of questions was conducted and data was collected from 30 lecturers who already have experience with online teaching during the lockdown. Lecturers were not selected randomly, they were rather selected based on their experience with online learning. Where 96% had greater than 5 years of experience, while only 4% had below 5 years of experience. Various programs such as foundation/diploma with 26%, undergraduate with 48%, and postgraduates with 26% were involved, where lecturers were classified according to their teaching skills and competencies. The majority of lecturers had intermediate level experience with 82%, followed by advanced skills with 11% and lastly beginner skills with 7%.

Since they used a scale of 5 which is strongly agree, agree, undecided, disagree, and totally disagree, as the perception of lecturers towards online teaching, they found that 80% are satisfied with online teaching and its approach, 63% are stressed with online teaching, and 70% says that the preparation for online teaching takes a lot of time.

However, this study focused more on undergraduate programs with 48%, if other programs were given an equal percentage the results of this study would change.

### 2.2.3. Benefit of online teaching and learning

* Vaibhav Gallani (2020) conducted a study to investigate the effectiveness of online learning during COVID-19. This study tries to understand various education difficulties such as connectivity problems, data consumption, flexibility and other factors relating to online teaching from the view of students and further concludes whether online learning can be used as a replacement for contact learning. This study focused on students who have at least 2 months experience of attending online.

This study used a questionnaire as a tool to collect data through google forms, they went through the different opinions and made a proper revisions as well as making necessary modifications. A sample of 170 students was included, where 62 students were females and the remaining 108 students were males. Furthermore, a sample of 121 students equivalent to 71.18% were undergraduate students, and a sample of 49 equivalent to 28.82% students were postgraduate students. About 2.35% of the collected data fell within the age of 15-18 years, 87.65% fell within the age of 18-22years and the other 10% fell within above 22years.

In their results they found that from the undergraduate program in the bachelor’s (B.com, BBA, BCA, BSC, BA) 17 students were agreeing with the quality of online learning, 35 students were disagreeing, 30 students were neither agreeing nor disagreeing, 6 students were strongly agreeing and 33 students were strongly disagreeing, while from the postgraduate level (M.Com, MBA, MCA,MSC,MA), 7 students were agreeing to the quality of online education, 20 students were disagreeing, 8 students were neither agreeing nor disagreeing, 0 students were strongly agreeing and 14 students were strongly disagreeing. Overall 56.45% of females and 62.04% of males believe that online teaching is ineffective and therefore cannot be used to replace contact teaching, while 43.55% of females and 37.96% of males believe that online teaching is effective enough to replace contact teaching. Therefore, it can be said that most students concur that face-to-face instruction cannot be replaced by online instruction. Students were found not to have the energy and enthusiasm to attend online teaching, and also the consumption of data during online classes was the main contributing factor.

However this study used a large number of undergraduate students as compared to postgraduate students, it is a fact that postgraduate students have more University experience than undergraduate students, therefore if an equal number of postgraduate students were considered the results of this study may change.

* Suprianto, et.al (2020) conducted a study to determine the effectiveness of online learning during the COVID-19 pandemic concerning reactions, one’s behavior, learning approach, and results obtained. During this period the campus learning platforms such as Whatsapp, Webex, and as well as google classroom were considered.

This study used both the qualitative and quantitative approaches to complete their research. With the quantitative method, a sample of 39 students was drawn, while for the qualitative approach 3 students were selected to be informants. A questionnaire, detailed interviews, and observations were used as tools to collect data.

In their results, they categorized their option into intervals, where interval 4-6 is not satisfactory in online learning, 7-9 is less satisfied with the online learning, 10-12 is just enough, 13-16is satisfactory, and 17-20 very satisfied. Now 5.13% of students were found to be very satisfied, 51.28% were satisfied with the implemented online learning, 33.33% were enough, 7.69% were less satisfied and 2.56% were not satisfied. Overall it was found that the majority of students are satisfied with online learning and think it’s effective. Consequently there are quite a number of factors that can influence the effectiveness of online learning, however, this study mentioned a few, and it does not detail how these factors will affect the current online learning approach.

* Awal Bahasoan, et.al (2020) conducted a study to determine the effectiveness of online learning during the COVID-19 pandemic. This study used the conditions under which online learning was carried out and also the cost of online learning as compared to contact learning as factors that affect the effectiveness of online teaching/learning in the COVID-19 pandemic.

In order to finish this study, which takes a quantitative approach, a survey was created. An active student from the management study program was selected as a sample of this study, using the simple random sample method. Online surveys were sent to 115 people in order to collect data, they further analyzed the collected data to describe it.

In their results, the factors affecting the effectiveness of online were given the 5 scales, very satisfied, quite satisfied, satisfied, less satisfied, and unsatisfied. Overall, it was discovered that online education during the COVID-19 epidemic is both efficient and effective. It is efficient because it is less expensive than in-person learning, but effective because of the laws requiring online education.

* I Made Satyawan, et.al (2021) conducted a study to describe the effectiveness of online learning using Undiksha e-learning in the physical education  and health study program during the COVID-19 pandemic. Factors such as the level of understanding of the student, materials for online learning, and students' motivation in taking part in the online learning were considered to be influencing the effectiveness of remote learning. Additionally, this study employs current contact learning to assess the efficacy of online learning.

This study uses a quantitative approach to complete the research. Now a sample of 120 students from the contact education method program was selected. A questionnaire was used as a tool to collect data, the validity of the content was then tested, construct the validity and check for the reliability of instruments, change the questionnaire tools into google forms, determining the research subject, distributing the google form links Therefore instruments were distributed as a google link for people to fill. The focus was on students who are doing Physical and Health study programs.

The results of this study showed that it is effective in Physical Education and Health study programs.

### 2.2.4. Ways to improve online teaching and learning

* Sutarto Sutarto, et.al (2020) conducted a study to understand and explore the strategies used by teachers of the SDIT Rabbi Radhiyya Curup to increase student’s interest in learning and to express student’s responses in online learning during the COVID-19 pandemic.

The study used a qualitative approach. Students, teachers, the vice principal of the curriculum, and the principal all served as informants. A semi-structured interview was conducted as a tool to collect data, the conducted interviews were analyzed using the Miles and Huberman model.

The study's conclusions demonstrated that teachers could engage students' interest by using a variety of techniques, including making learning materials engaging, succinct, and very clear in order to use easy and interesting media, conducting daily and repeated evaluations, and teaching students the value of learning.

Although this study showed that some students say online learning is fun because they can watch recorded sessions at their most convenient times, it has also shown that students lack the ability to be together with their friends.

* Deli Reyan Rizaldi (2021) conducted a study to evaluate the strategies to improve teacher ability in using the madrasah e-learning application during the COVID-19 pandemic. This study has outlined some applications which are expected to enable the increase in students' motivation as well as enthusiasm for learning online from home during this pandemic and still be able to apply health protocol in order to minimize the spread of the virus.

This study has selected 36 teachers who teach Mardrash Aliyah Plus Nurul Islam as participants. In order to complete this research, 3 stages were followed, one being the preparation stage, two which was the implementation stage, and the last stage which was the evaluation stage, where the assessment process is conducted to secure the original objectives of activities achieved.

The results of this study showed that teachers need mentoring activities in order to understand the utilization of ministry of religion online learning applications with different supporting features.

## 2.3. Chapter Summary

The study reviews in this chapter support the hypothesis that how effective online teaching and learning is depends on various factors such as necessary skills and resources, understanding the pedagogical approaches suitable for the online environment, and socio-economic factors (Silvana di Gregorio, 2020).

# CHAPTER 3 Research Methodology

## 3.1. Introduction

This chapter aims to test the effectiveness of online teaching/learning (Kaushal Kumar Bhagat, 2021). The methodology will be followed to accomplish the objectives of this study.

To achieve the aim of this study, Twitter opinions will be used as a data set, this will help in knowing what various people in South Africa think of the current online learning.

## 3.2. Data collection

Data will be collected from Twitter, because Twitter users will be used to gather data since they are helpful for current real-world circumstances like the COVID-19 pandemic issue, another reason is that Twitter has gained popularity as a social media platform with a sizable and quickly expanding user base where users post status updates known as tweets that convey their opinions about their experiences using particular products, services, and emerging trends (Arambepolas, 2020).

The objective is to accept a search word term from a user, find the current sentiment for that term on Twitter, and save the most recent 100 tweets that match that search keyword then use a machine learning classifier to categorize each of these tweets as positive, negative, or neutral.

In order to access data from Twitter, we need to access the Twitter API using the Python-twitter module, but first register your application on Twitter and generate an API key and credentials on <https://aps.twitter.com>(Makice, 2009)

## 3.3. Data Cleaning

Because there are irrelevant elements in tweets, it is necessary to perform data pre-processing on tweets, this is to standardize specific tokens of tweets. Data standardization is transforming data into a standard format so that it may be processed and analyzed by others.

The performance of data pre-processing includes transforming the raw dataset into an understandable format, to remove unwanted noise from the textual data using steps such as;

NTLK and Regular Expressions (REs) will be used to perform data cleaning. REs are a sequence of characters that define search patterns, they are used to extract patterns from tweets, meaning REs can be used to find all words that match the pattern (Christopher M, 2019) using quantifiers. RE is also used to:

* Convert all texts into lower case
* Replace links with the string ‘URL’
* Replace @ with ‘at\_user’

And to;

* Remove all the hashtags, user mentions, and emoji
* Remove stopwords (including URL and user)
* Remove all the punctuations.
* Remove all the numbers.
* Remove parentheses

We will use NLTK

## 3.4. Model training and testing

Data that will be used for training will be downloaded from a corpus. After pre-processing tweets, extract features from both the training and test sets of data. The opinion mining approach is a classification problem that classifies the document as positive or negative. For this Sentiment Analysis, we will use Naive Bayes (NB) to compute posterior probabilities (M.S. Farooq, 2011).

P (document is positive|Words)

=

P (document is negative|Words)

=

Therefore the class whose posterior probability is greater will be picked.

To determine the polarity of a word we use a lexicon, such as a dictionary (Marr, 2018), a lexicon is a resource with information about words and has information such as lists of words that are positive and negative.

To classify tweets using the Naive Bayes classifier first, build a vocabulary (list of all the words in all the tweets in the training data), and represent each tweet with the presence/absence of these words in the tweets, then use NLTK’s built-in NB classifier to train the classifier. To conclude the sentiment of the word, take the majority vote and the percentage of tweets with that sentiment and print the output.

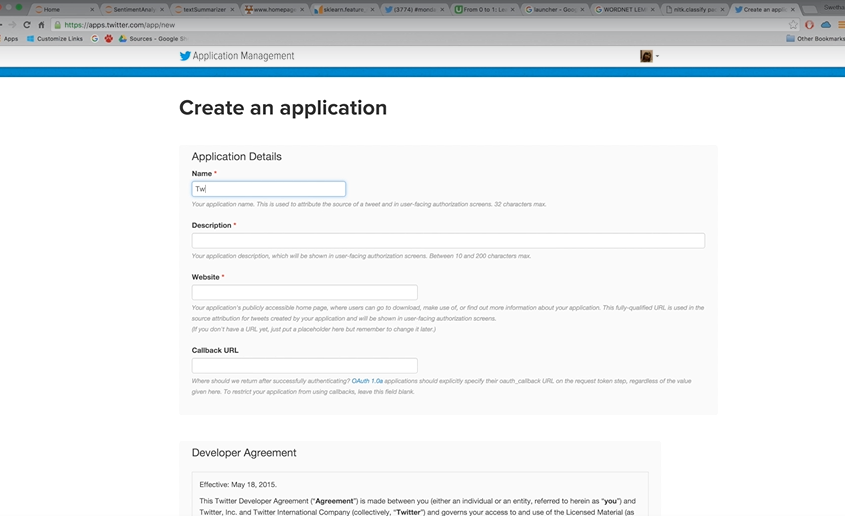
## 3.5. Data Visualization

Data visualization is the order that aims to interpret it and reveal patterns, trends, and connections that might not otherwise be visible. Python libraries such as matplotlib and seaborn will be deployed for data visualization. Matplotlib is a plotting python library that is frequently used in the Python community (Sunday O.A., 2012), and can also be used in the Jupyter notebook. Therefore matplotlib will be used to create charts plots, bar charts, pie charts, histograms, and others (DataCite, 2019). Based on Matplotlib, the NumPy and pandas data structures are closely linked with the Python data visualization package Seaborn. (Ramesh, 2013). It internally carries out the required statistical aggregation and mapping operations to produce the desired user-requested informative visualizations. To provide the desired user-requested informative visualizations, it internally performs the necessary statistical aggregation and mapping processes.

## 3.6. Techniques to evaluate

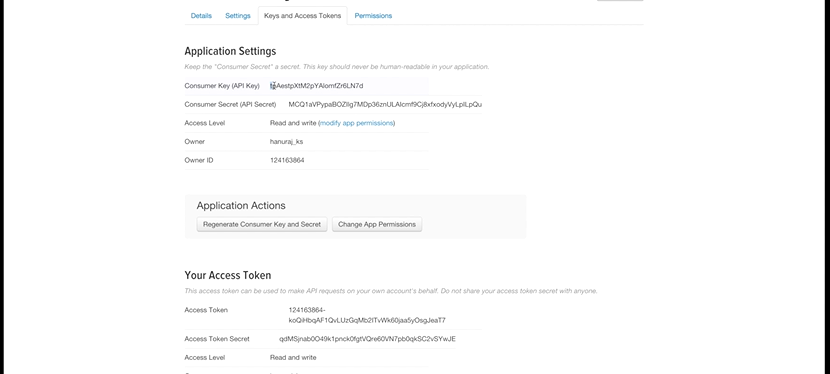
### 3.6. 1.Twitter API Application

The first step will be creating a Twitter developers account in order to get the API and Access keys. Below shows a screenshot of a Twitter API page for creating an application.



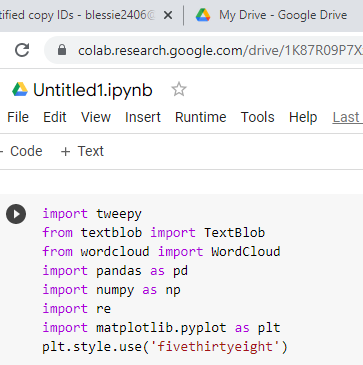
### 3.6.2. Twitter API keys

After successfully creating an account, you will get your API/Access keys, from the Twitter developer page.



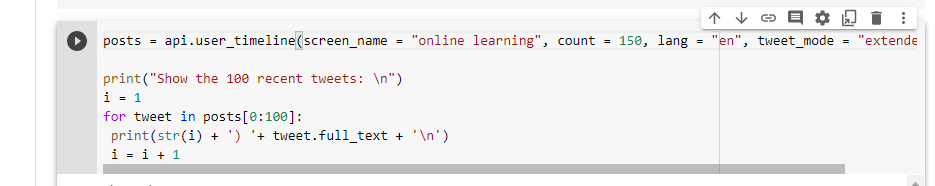
### 3.6.3. Importing Libraries

Import libraries that are needed, then copy your API keys and access keys and paste them into the code then verify them.



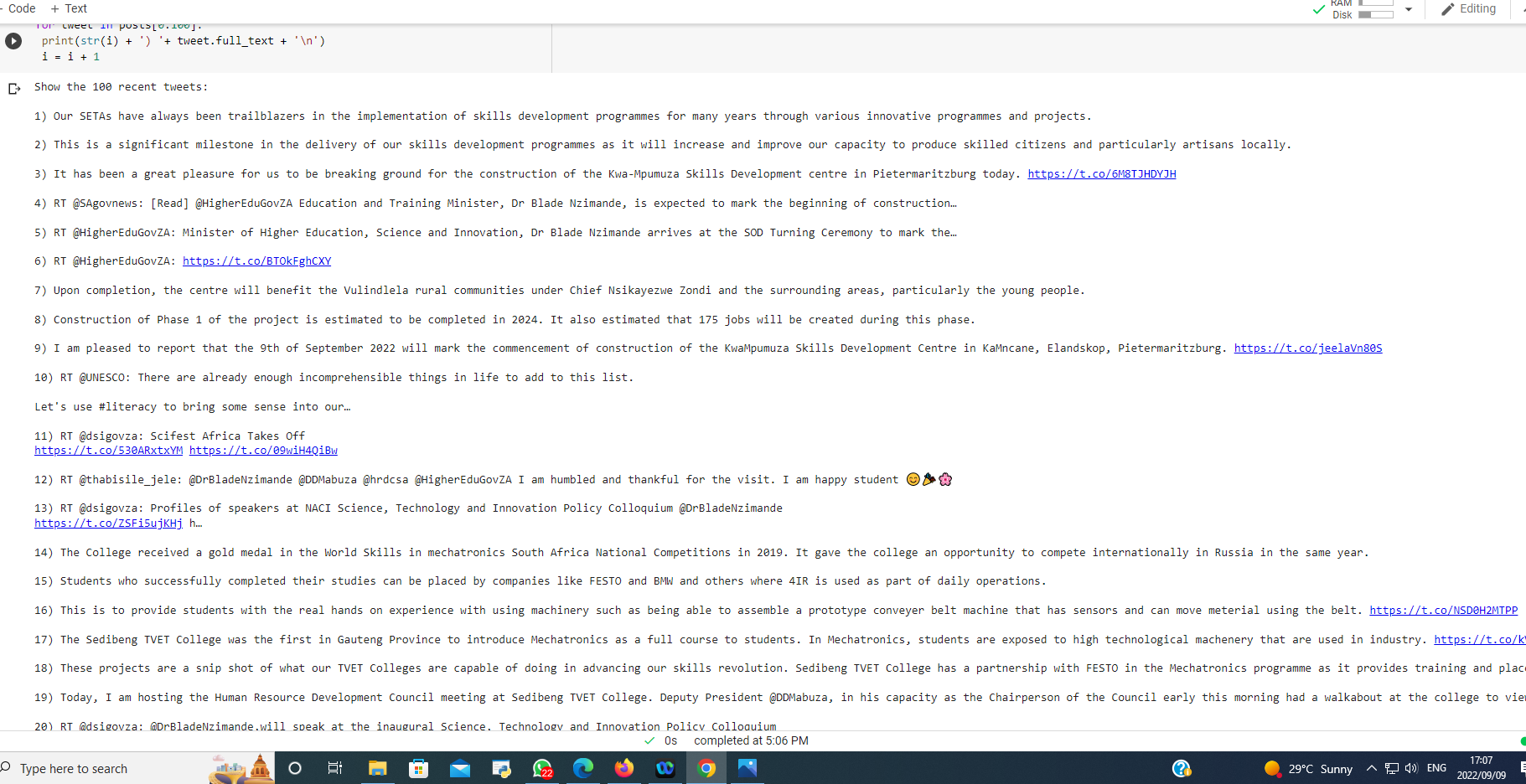
### 3.6.4. Search a Term

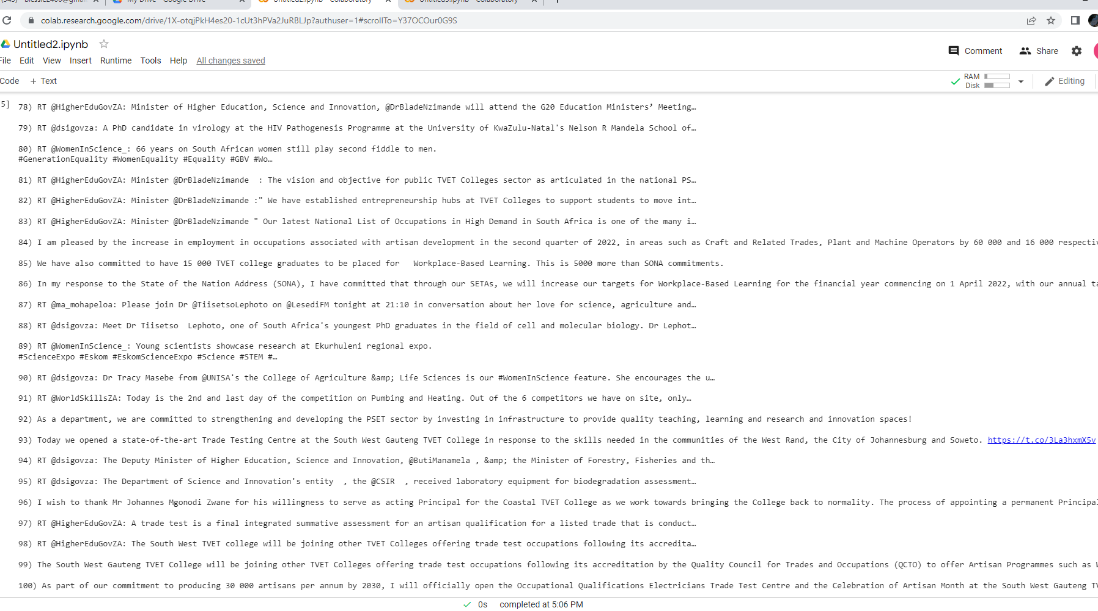
The below mini code allows you to search any term or any Twitter user account. After inserting your desired term this code will run and return a collection of tweets and here I have chosen a minimum of 100 tweets at once.



### 3.6.5. Initial Dataset

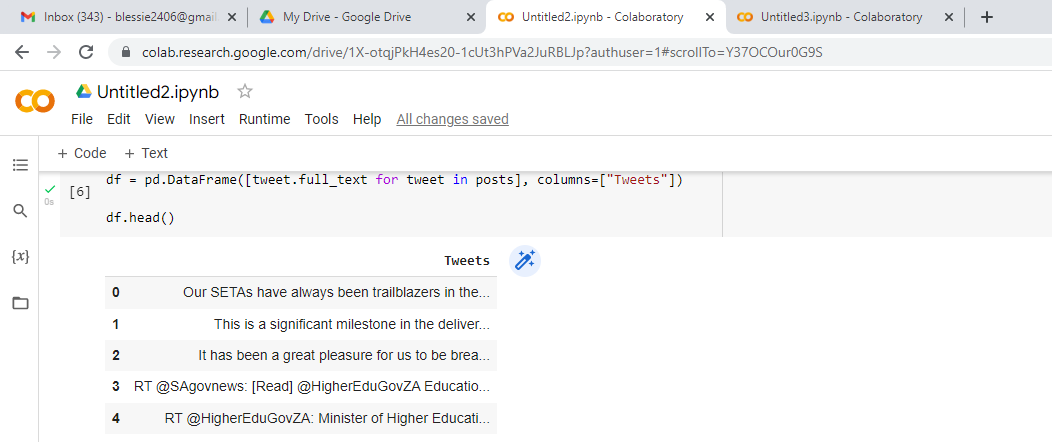
The initial dataset for this study is a set of 100 tweets collected directly from Twitter using python colab code. These are the opinions of some Twitter users about online teaching/learning. These tweets do not show the name of the user who posted them.





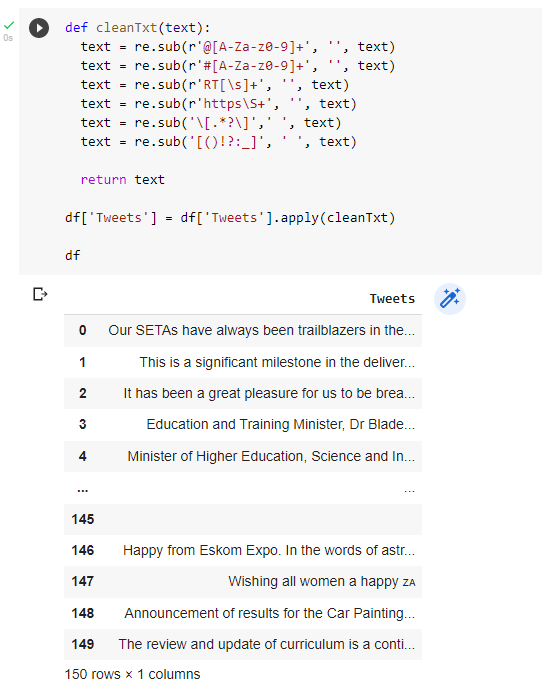
### 3.6.7. Data Frame

From the tweets collected, create a pandas dataframe of these tweets so that they can be analysed easily, and make cleaning simple. Below are the tweets that have been added into the dataframe, these tweets are not yet cleaned and still contain noise.



### 3.6.8. Data Cleaning and Preprocessing

Having data that is cleaned will primarily increase productivity and result in the highest quality of information in decision-taking. The image below shows a code to clean the dataset, and the final data set that is cleaned and free of all the noise. It is evident from the cleaned data set in the dataframe that special characters like @ have been removed.



# 3.7.1. Extract features and train a classifier

Since we are using the Naive Bayes method, we will use NLTK’s built-in classifier to perform the classification. First, build a vocabulary, this will give us a list in which all the words in all the tweets are present, these have to be de-duped. Each word appears in this list as many times as it appears in the corpus. Create a dictionary with each word and its frequency and return the unique list of words in the corpus. NLTK has an apply\_features function that takes in a user-defined function to extract features from training data. We want to define our extract features function for each tweet in the training data and represent it with the presence or absence of a word in the vocabulary. This will give us a dictionary, with keys like contains word1 and contains word2. We can also extract the features and train the classifier. Apply features will take the extract features function defined above and apply it to each element of the training data. It automatically identifies that each of those elements is actually a tuple, so it takes the first element of the tuple to be the text and the second element to be the label and applies function only to the text. We will therefore have a classifier that has been trained using Naive Bayes.

## 3.8. Chapter Summary

This chapter explains in detail the methodology of the study. The codes in this study were all running on python google colab, with each important library imported.

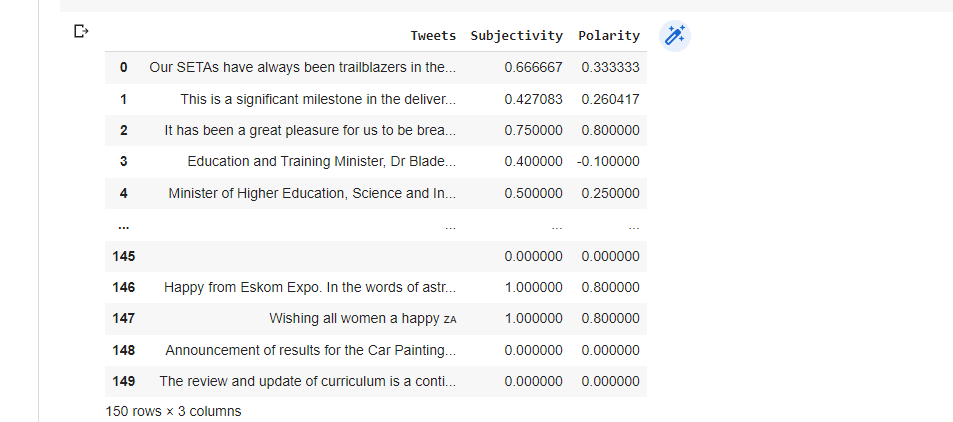
# Chapter 04 Results and Analysis

## 4.1. Introduction

The data that has been prepared and used for training evaluated to measure sentiment toward online learning is the basis upon which the sentiment analysis model is constructed. In this chapter, the results are analysed and visualised. The results obtained were obtained independently using different pieces of codes. Matplotlib and seaborn were deployed to get data visualisation.    `

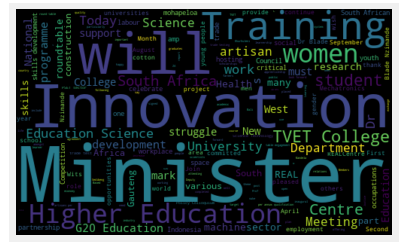
## 4.2. Polarity and Subjectivity

Examining a body of text in order to completely grasp the opinion it communicates is the basic objective of sentiment analysis. Typically, we measure this feeling with a polarity which is a positive or negative value. Polarity is a float, which falls between [-1, 1] where -1 is said to have a negative polarity and 1 to have a positive polarity. Subjectivity determines how much real and personal information is present in the text. For this study, subjectivity and Polarity were found to have the following results.



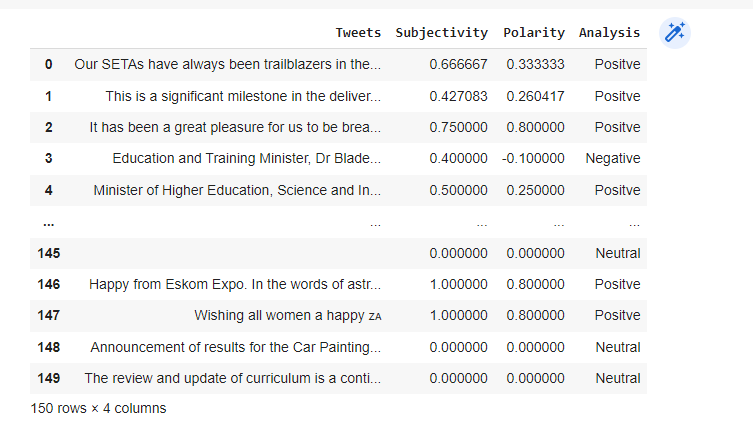
## 4.3. WordCloud visualisation

After the pre-processing steps of tokenization and data conversion to remove extra noise, the results were sorted to obtain the highest word frequency before being visualized into positive and negative word clouds, as seen in the image below. The size of each word indicates the relative frequency of that word, meaning a word appears in the word cloud bigger and bolder the more times it appears in the text.



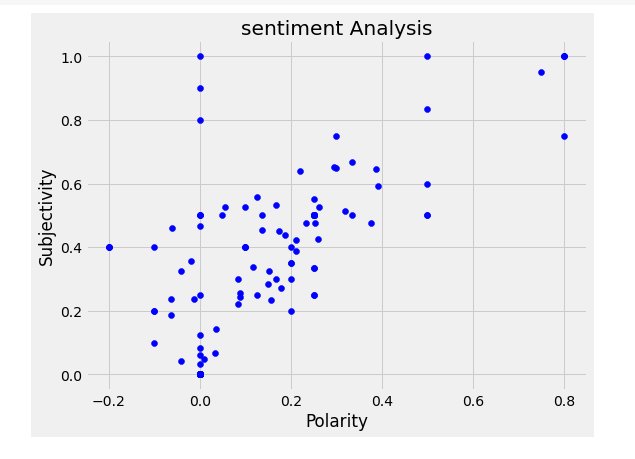
## 4.4. Sentiment Analysis model

Due to training the data that has been prepared and evaluated to assess sentiment toward remote learning, a sentiment analysis model is created. From the measured Polarity and Subjectivity, we found the analysis, which specifies whether a specific tweet is negative, positive, or neutral. The figure below shows those sentiments specified.



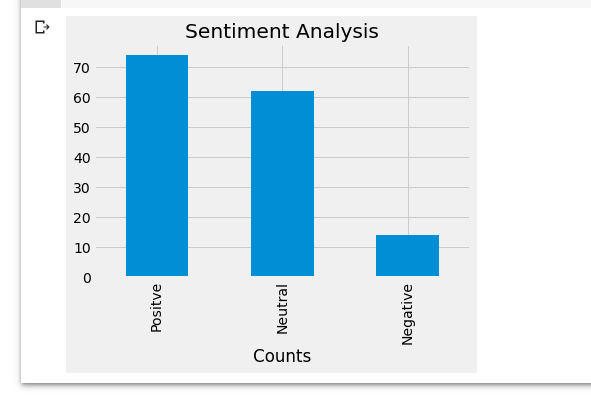
## 4.5. Distribution Plot

The findings of this plot shows the Sentiment Analysis distribution. Based on this distribution, most tweets are lying on the Positive side, while less tweets are lying on the negative side. It was found that many students perform better in online teaching/learning. These results also show that more lecturers have adopted the online learning method so well.



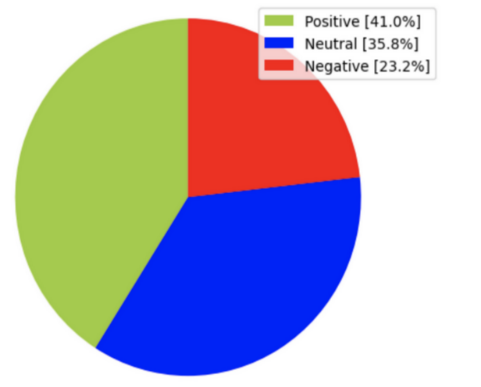
## 4.6. Bar Graph results

According to this graph's sentiment measurement data, more than 70 percent of tweets were positive, more than 10 percent were neutral, and the remaining tweets with a sentiment score of 60 or higher were deemed neutral. According to these results online learning has been moderately used in South Africa throughout the COVID-19 Pandemic, as seen from the slight difference between the positive tweets and the neutral tweets. These results show that even with all the challenges students face during online learning, they are still able to perform well regardless.



## 4.7. Pie chart

The results further showed that from the dataset collected 41% of the Twitter community believes that online learning is effective, while 23.2% think online teaching/learning is not at all effective and the remaining 35.8% are neutral.



## 4.8. Objective 2

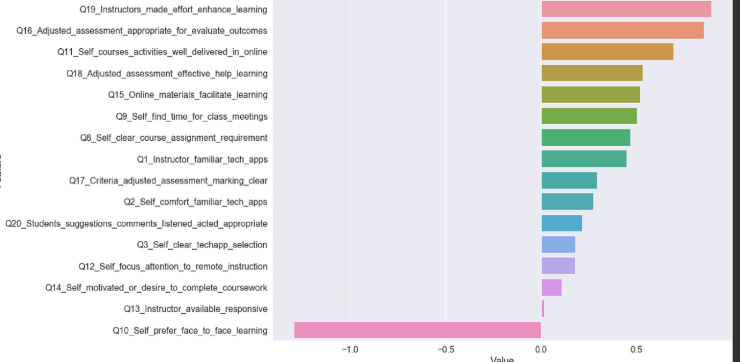
### 4.8.1. Factors that affect the effectiveness of online teaching/learning

* **Readiness**- How comfortable students and instructors are with the required technologies or applications for online learning, or rather the difficulties they face accessing reliable communication tools.
* **Accessibility-** How often specialized tools, such as libraries, and live-streaming lectures be accessed for studying and teaching.
* **Instructor’s related-** How comfortable or familiar instructors are with the required technologies or applications for this online teaching/learning method.
* **Assessment related-** How clear are the assessments to students
* **Learning related-** How well are the course lessons delivered in the online environment.
* **Self-concern-** Howconcerned students are about their grades, and how concerned the instructors are about student’s performance.

## 4.9. Objective 3

**4.9.1.** This objective was from one of the studies I have reviewed by Indy Man Kit Ho, they collected the Twitter dataset to predict the satisfaction of students’ performance during COVID-19 using the factors that they believed were contributing to the student’s performance. In their research, they classified the tweets and used the Naïve Bayes classifier to compute the probabilities of these factors in order to determine how much they were contributing to students’ performance.

Their results are shown below which show that the instructor’s availability has a lower probability meaning that whether the instructor’s response time is fast or slow, it is still the responsibility of students to make sure that they study hard. In comparison, the instructor's efforts have a higher probability, meaning that instructors’ efficient efforts in different programs to enhance students’ learning contribute higher to the student’s performance.



## 4.10. Chapter Summary

After performing all the objectives, the results obtained show that a high percentage of Twitter opinions believe that online teaching/learning is effective, and all the factors mentioned in their probabilities were according to how they positively affect the student’s performance.

# Chapter 05 Conclusion and Feature Work

## 5.1. Introduction

This chapter intends to discuss the summary of the research findings. These are the results of remote teaching and learning and how it has affected the people in South Africa. This chapter will refer to all the objectives that have been achieved, it will also show future work, conclusions, and as well as limitations.

## 5.2. Answers per objective

### 5.2.1. Objective 1

From objective 1 which was to perform the sentiment Analysis, the results have clearly shown that according to Twitter opinions a large number of people think online learning is the best method, this was due to the performance of students during online learning and during contact learning. The results show that.

### 5.2.2. Objective 2

There is quite a number of factors that can affect student performance, in this paper we only mentioned a few that we believed were contributing top to the student’s performance. The listed factors show how they each can influence the student’s and instructors’ performance during remote teaching and learning.

### 5.2.3. Objective 3

This objective outlines the probability of satisfaction with regard to each factor in students’ performance. It shows how each factor does justice to online learning, and how satisfied the people on Twitter are with regard to each factor. Finally, it shows that instructors’ availability does not have a high probability meaning students still hold the responsibility to perform better, whether the instructor is there or not. And further shows that the instructor’s efforts hold a higher probability meaning the amount of effort the instructors put into making the remote teaching and learning work it highly influences the student’s performance.

## 5.3. Future Work

Students are different and come from different backgrounds and they also have different views based on different factors therefore more factors can be added to the ones investigated to further check the satisfaction of students’ performance. A sentiment Analysis depends on the dataset collected, a different dataset can be collected from a different site like Kaggle to get different results. Other machine learning algorithms such as the random forest can be used to predict the satisfaction of students’ performance.

## 5.4. Limitations

This study only focused on testing the effectiveness of online teaching/learning based on Twitter opinions, and the factors that influence this method. Each factor was found to contribute a certain probability. Therefore in this study, the dataset was strictly based on Twitter sentiments, not the combined datasets that might include other sites as well.

## 5.5. Chapter Summary

According to factors that influence the satisfaction of students’ performance, there could be measures taken to improve the existing online learning, that will result in the percentage of people who believe online teaching/learning is effective rising, reduce the percentage of neutral sentiments, and as well as reducing the percentage of negative sentiments

# References

* Anon., 2020. *What is remote learning and how effective can it be?.* s.l.:s.n.
* Arambepola, N., 2020. *Analysing the Tweets about Distance Learning during.* Kelaniya: s.n.
* Arambepola, N., 2020. *Analysing the Tweets about Distance Learning during.* Kelaniya: s.n.
* Arambepolas, N., 2020. *Analysing the Tweets about Distance Learning during COVID-19.* Kelaniya: s.n.
* Chu, K., 2021. *Text Processing in Python.* s.l.:s.n.
* Di Pietro, G. B. F. C. P., 2020. *The likely impact of COVID-19 on education.* Europe: http://www.stock.adobe.com/.
* Di Pietro, G. B. F. C. P., 2020. *The likely impact of COVID-19 on education.* European Commision: s.n.
* Ghebreyesus, D. T. A., 2020. *WHO Director-General's opening remarks at the media briefing on COVID-19.* s.l.:s.n.
* Kaushal Kumar Bhagat, S. M. A. D. a. C.-Y. C., 2021. *Public Opinions about Online Learning during COVID-19.* Switzerland: MDPI.
* Khandewal, H., 2020. *Sentiment Analysis of a Tweet With Naive Bayes.* s.l.:s.n.
* Özüdoğru\*, G., 2021. *Problems faced in distance education during Covid-19 Pandemic.* Turkey: s.n.
* Reimers, F., 2020. *Supporting the continuation of teaching and learning during COVID-19 Pandemic.* s.l.:s.n.
* Syafrida Hafni Sahir1, R. S. A. R. M. F., 2021. *Online learning sentiment analysis during the covid-19.* Indonesia: s.n.
* Wu, D. G. M. L. A. C. A., 2020. *The impact of COVID-19 (Coronavirus) on global poverty: Why Sub-Saharan Africa might be the region hardest hit.* s.l.:Data Blog.
* Anon., n.d. *Sentiment Analysis: Comprehensive Beginners Guide.* s.l.:https://getthematic.com/sentiment-analysis.
* Arambepola, N., 2020. *Analysing the Tweets about Distance Learning during COVID-19 Pandemic using Sentiment Analysis.* Kelaniya: s.n.
* Kaushal Kumar Bhagat, S. M. A. D. a. C.-Y. C., 2021. *Public Opinions about Online Learning during COVID-19: A Sentiment Analysis Approach.* India: s.n.
* Silvana di Gregorio, 2020. *USING NVIVO FOR YOUR LITERATURE REVIEW.* London and Boston: s.n.
* Thi Tinh Thuong Pham, H. A. L. a. D. T. D., 2021. *The Factors Affecting Students’ Online Learning Outcomes during the COVID-19 Pandemic: A Bayesian Exploratory Factor Analysis.* Ehsan Namaziandost ed. s.l.:Hindawi.
* Ho, I. M. K., 2021. *Predecting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques.* Hong Kong: s.n.
* Shim TEE, L. S., 2020. *College students experience of emergency remote teaching due to COVID-19.* South Korea: s.n.
* Sun L, T. Y. Z. W., 2020. *Coronavirus pushes education online.* South East: s.n.

# Appendix

For Sentiment Analysis

# Description: This is a program that shows Twitter API Keys for my account, this are the keys that enables the user to be able to fetch any specified tweets from Twitter.

consumerKey = "If0yHI0ScqAQbKoSm2QC7mcLL"

consumerSecret = "FAlLMvWVmwunroYcGnpv5ol9jBUBWp3OuzMM8bdCRFTNc7rPjq"

accessToken = "1258797933195079680-DCYFHMtBsj0NOgvn77F8Jf55QdCLj6"

accessTokenSecret = "qa6t0l8JszKg6R0S9ReW2QVPxrvLV7IAE2NguLruG7inD"

authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)

#Program to verify Twitter credentials(API Keys)

authenticate.set\_access\_token(accessToken, accessTokenSecret)

api = tweepy.API(authenticate, wait\_on\_rate\_limit=True

#program that indicates how to search for a term on Twitter

posts = api.user\_timeline(screen\_name = "online learning", count = 150, lang = "en", tweet\_mode = "extended" )

print("Show the 100 recent tweets: \n")

i=1

for tweet in posts[0:100]:

  print(str(i) + ') '+ tweet.full\_text + '\n')

  i = i + 1

#program that indicates how to insert the dataset into a dataframe

df = pd.DataFrame([tweet.full\_text for tweet in posts], columns=["Tweets"])

#program that indicates how to get subjectivity and polarity

df.head()

def getSubjectivity(text):

  return TextBlob(text).sentiment.subjectivity

def getPolarity(text):

  return TextBlob(text).sentiment.polarity

df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)

df['Polarity'] = df['Tweets'].apply(getPolarity)

df

#program that indicates how to get the wordcloud

allWords = ' '.join([twts for twts in df['Tweets']])

wordCloud = WordCloud(width = 500, height = 300, random\_state = 21, max\_font\_size = 119).generate(allWords)

plt.imshow(wordCloud, interpolation = "bilinear")

plt.axis('off')

plt.show()

#program that indicates how to get the sentiment Analysis from subjectivity and polarity

def getAnalysis(score):

  if score <0:

    return 'Negative'

  elif score == 0:

      return 'Neutral'

  else:

        return 'Positve'

df['Analysis'] = df['Polarity'].apply(getAnalysis)

df

#This code outputs all the positive tweets from the collected tweets

j=1

sortedDF = df.sort\_values(by=['Polarity'], ascending = 'True' )

for i in range(0, sortedDF.shape[0]):

  if(sortedDF['Analysis'][i] == 'Positive'):

    print(str(j) + ') '+ sortedDF['Tweets'][i])

    print()

    j = j+1

#This code outputs all the negative tweets from the collected tweets

j=1

sortedDF = df.sort\_values(by=['Polarity'], ascending = 'False')

for i in range(0, sortedDF.shape[0]):

  if(sortedDF['Analysis'][i] == 'Negative'):

    print(str(j) + ') '+ sortedDF['Tweets'][i])

    print()

    j = j+1

#This code indicates how to get the sentiment Analysis plot with subjectivity vs polarity

plt.figure(figsize=(8,6))

for i in range(0, df.shape[0]):

  plt.scatter(df['Polarity'][i], df['Subjectivity'][i], color = 'Blue')

plt.title('sentiment Analysis')

plt.xlabel('Polarity')

plt.ylabel('Subjectivity')

plt.show()

#This code indicates how to get the bar graph

df['Analysis'].value\_counts()

plt.title('Sentiment Analysis')

plt.xlabel('Counts')

df['Analysis'].value\_counts().plot(kind= 'bar')

plt.show()