SENTIMENT ANALYSIS ON REMOTE LEARNING IN KENYA.

# **Business Overview**

## Overview

The rise in online learning in recent years has been accelerated by technological advancements.Recently, COVID-19 has resulted in many schools shut down across the world.The UNESCO reports that learning disruption affected over 1.2 billion learners globally,with the distinctive rise of e-learning ,whereby teaching is undertaken remotely and on digital platforms. School closures precipitated a massive unprecedented shift in how

instructional design and delivery is handled. Overnight, we saw schools switch to remote learning in order to mitigate the effects of the disruptions to the learners.

With the sudden shift away from the classroom in many parts of the globe,some are wondering whether the adoption of online learning will continue to persist post-pandemic ,and how such a shift would impact the worldwide education market.

## **Problem statement**

As a result of covid, most education systems were forced to resort to remote learning fully, some partially and some relented to the idea due to uncertainty of how their students will take it up and the effectiveness of remote learning. We decided to look into people’s sentiments about it and understand their perceptions so as to narrow down the uncertainties and maybe tell whether it could be the new normal even post the pandemic. This could be used by educational institutions in their decision making concerning how to incorporate remote learning into their systems.

## 

## **Objective**

The main objective for this study will be to explore different perspectives on remote learning given the increasing use and adoption of remote learning and its chances of being adopted fully as a learning system.

### **Specific Objectives**

1. To determine whether remote learning is prominent in all learning institutions.
2. To determine the most preferred form of learning.
3. To perform sentiment analysis on each tweet in our dataset and classify whether it is positive or negative

### **Research Questions**

1. Is remote learning likely to be adopted fully by all learning institutions?
2. What is the most preferred form of learning(remote learning vs traditional learning)?
3. What sentiments are dominant ? Are they positive or negative?

## Impact

This research seeks to influence decision making in Educational boards regarding remote learning and what approach they are going to use incorporating it in their systems. It will also be of help to them to understand the various associated members' perceptions as they make decisions for future purposes.

## Data Mining Goals

Our data mining goal for this project is to investigate the perceptions and perspectives of Kenyans on twitter concerning remote learning.Potential questions for consideration include:

1. Why is remote learning mostly preferred or not preferred?
2. What does the future look like regarding remote learning?

## Success Criteria

A successful analysis will provide insights on the perceptions of remote learning and be able to predict how remote learning will dominate post covid.

## Assessing the situation

Assumptions

The data provided is accurate.

Resource inventory

Twitter datasets :

1. Softwares:

* Github [<https://github.com/iyline-sigey/PREDICTIVE-ANALYSIS-ON-REMOTE-LEARNING>]
* Google collab for project analysis
* Twitter(developers account) for data scraping
* Tweepy api,
* Streamlit/plotly dash
* Tensorflow/keras
* Jira for project tracking [<https://lynnechumo.atlassian.net/jira/software/projects/G1M2P/boards/2>]

**Implementation plan**

| Phase | Time-Frame |
| --- | --- |
| Formulation of Research Question | 30 minutes |
| Business Understanding | 1 hour |
| Data Understanding | 1 hour |
| Data Preparation and Cleaning | 5 hours |
| Data Analysis | 1 hours |
| Summary and Conclusion | 30 mins |

# **Data Description**

For this analysis, the data available has 5 columns and 93 rows collected from 2018 to 2021. The information available in the dataset is described in the table below:

| COLUMN | DESCRIPTION |
| --- | --- |
| USER | Person who made the tweet |
| TWEET | The sentiments from the user |
| LOCATION | Location of the user |
| CREATED-AT | The time user’s account was created |
| RETWEET-COUNT | The number of times the sentiment has been retweeted |
| CLEAN-TWEET | The tweet after cleaning |

# **Data Preparation**

* **Scraping the Data**

We first scraped data from twitter

* **Loading and Reading the Datasets**

We first imported the required libraries to help in the analysis. After, we loaded our dataset and created a dataframe. We then previewed both head( first 5 rows) and tail( last 5 rows) to get an understanding of the information we shall be analysing.

* **Exploring the Data**

Afterwards, we checked the shape of our dataset. Our data contains 10 columns and 197 rows. We checked for the information of each column, to understand the datatypes of each.

* **Data cleaning**

1. **Uniformity**: We checked for the uniformity in naming the columns names and removed any white spaces in the our columns as well as renaming our column names to lowercase.
2. **Missing Values**: checked for any null values in our data. We had 16 missing values columns which we went ahead and dropped.
3. **Duplicates**: We checked if there were any duplicates in our dataset. Our data frame had a significant amount of duplicates(88) .We decided to drop them as they would affect the accuracy of our models.
4. **Irrelevant Columns**:5 Columns were irrelevant for our analysis, hence they were dropped.

* **Text Cleaning**

We also cleaned the tweets by removing the unnecessary words like emojis,punctuations,digits,urls and stopwords.We then performed stemming to remove the last characters in words that lead to incorrect meaning and spelling.

* **Feature Engineering**

We extracted data from the Sentiment column and stored them in a new column to specify if the tweet was either negative,positive or neutral to help us in visualizing the tweets.

* **Text Augmentation**

After feature engineering was done,we performed data augmentation to increase the size of our training data by generating more data from the data we had.

# Analysis

After cleaning our dataset, it was time to do analysis to answer our research objective. Detailed Analysis were conducted in google colab notebook and can be found in our github repository[<https://github.com/iyline-sigey/PREDICTIVE-ANALYSIS-ON-REMOTE-LEARNING>]

# 

# **Modelling**

**Data Preprocessing**

Machine learning algorithms need some sort of numerical feature vector in order to perform the task. Some of the basic text pre-processing procedures such as:

* Tokenization: converting our text into a list of tokens
* Stemming: which reduced inflected words to their stem, base or root form for e.g learning to learn
* Transformation of data by finding and plotting the most common words
* Extracting features such as urls, retweets, user information and description, conversion to lowercase and removal of stopwords
* Our dataset had a class imbalance which we solved by increasing the minority class through upsampling

**Models**

After our pre-modeling stages we then implemented the following algorithms to classify our tweets:

* Random Forest
* KNearest Neighbors
* Gradient Boost
* Naive Bayes
* Neural Networks: MultiPerceptron(MLP,

ANN and RNN

* Bert

**Model Performance**

| Model | Accuracy | Precision | Recall | f1-score |
| --- | --- | --- | --- | --- |
| Random Forest Classifier | 93.54 | 95% | 86% | 90% |
| MultiPerceptronClassifier | 66.12% | 61% | 63% | 62% |
| KNN Classifier | 71% | 68% | 69% | 68% |
| Gradient Boost Classifier | 69% | 69% | 71% | 68% |
| Naive Bayes | 63% | 33% | 48% | 39% |
| ANN | 60.75% |  |  |  |
| RNN | 80.30% |  |  |  |
| LSTM | 94.95 |  |  |  |

**Our Keras model was the best performing one with an accuracy score of 94.95% and loss of 0.2669 (26%)**

# Deployment

The models were then deployed by the use of Streamlit. We also used wandb to track our algorithms and parameters

**Conclusion**

Most of our tweets carried a positive sentiment which means that generally online learning was well received by Kenyans on twitter

Most of the institutions tweeting about online classes were higher learning institutions

# Recommendations

* We recommend that other learning institutions should also adopt remote learning as it is the most preferred form of learning during the current pandemic.
* Most of our institutions, especially educational ones, were built largely for the generations with no intuition of the digital age approaching and taking over. However in the advent of unprecedented circumstances such as the Covid-19 pandemic, the education sector and institutions should consider making adjustments in how they offer such services and conceptualize innovative ways to avail education to learners. Some of the strategies may include homeschooling, radio and television education programs, remote learning, online learning, distance learning, blended learning to make sure learning continues.