**MAKERERE UNIVERSITY**

**SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY**

**CONCEPT PAPER FOR RECESS FINAL PROJECT**

A Concept Paper submitted to the School of Computing and Informatics Technology

For the Study Leading to a Project Proposal in Partial Fulfillment of the

Requirements for the Award of the Degree of Bachelor of

Science in Software Engineering of Makerere University.

|  |  |  |
| --- | --- | --- |
| NAME | REG NO | STD. NO |
| KYAKA HERMAN CEASER | **15/U/6851/PS** | **215011936** |
| MALITI EMMANUEL | **17/U/5926/EVE** | **217006711** |
| KABWAMA RONALD | **15/U/5398/EVE** | **215018454** |
| KAIJUKA DONALD | **13/U/5658/PS** | **213004661** |

**Table Of Contents**

Contents

[**1** **Introduction** 3](#_Toc13841920)

[**1.1** **Title** 3](#_Toc13841921)

[**1.2** **Background.** 3](#_Toc13841922)

[**1.3** **Problem Statement.** 3](#_Toc13841923)

[**1.4** **Aim and Objectives** 3](#_Toc13841924)

[1.4.1 Aim or general objective 3](#_Toc13841925)

[1.4.2 Specific Objectives 3](#_Toc13841926)

[**2** **Methodology** 3](#_Toc13841927)

[**2.1** **Data collection** 3](#_Toc13841928)

[**2.2** **Data Analysis** 3](#_Toc13841929)

[**2.3** **Data visualizations and Graphical Analysis.** 3](#_Toc13841930)

[**2.4** **Predictive Modelling** 4](#_Toc13841931)

[**3** **References.** 4](#_Toc13841932)

# **Introduction**

## **Title**

Review prediction from text comment

**Definitions.**

A review is a formal assessment of something with the intention of instituting change if necessary. Knowledge of how the commenting community reacts to a specified topic.

## **Background.**

Data set: Movie Reviews, <https://www.kaggle.com/nltkdata/movie-review>. This data was first used in Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proceedings of the ACL, 2004.

The structured dataset contains comments from a single website about a movie and it is identified as a text column. It contains comments by people from various locations or countries.

## **Problem Statement.**

There is a lot of time wasted in reading individual comments to award the review to the subject topic. This however can be automated by involving machine learning to generate the review.

## **Aim and Objectives**

### Aim or general objective

To Develop a Prediction Model that uses the existing datasets to predict likelihood of a good or bad review on the subject.

### Specific Objectives

* To collect and analyze datasets to be used in training our prediction model.
* To Identify the variables that are to be used in rating subjects and be able to predict the likelihood of good or bad review basing on those variables.
* To use the collected data to visualize the relationships between the chosen variables.
* To design a prediction algorithm that will be used to predict review.
* To implement the algorithm using Machine learning tools like Natural Language Processing Tool Kit, logistic regression and the sample datasets chosen.
* To test the accuracy of the implemented Predictive model.

# **Methodology**

## **Data collection**

This includes all the techniques used to collect data.

We shall use Kaggle, a dataset repository on the web to get sample datasets that satisfy the variables listed and download them to our local server for data mangling.

## **Data Analysis**

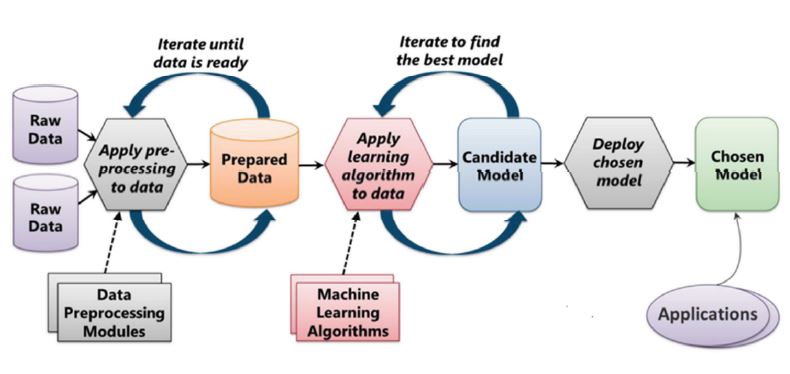
In this step, we shall use NLTK as a tool for analysis, whereby we shall use matplotlib, pandas and numpy libraries to clean and filter out relevant data from our datasets.

## **Data visualizations and Graphical Analysis.**

This step includes displaying the cleaned data in step [2.2](#_Data_Analysis) above.

## **Predictive Modelling**

In this step we shall have understood the relationships between the different variables. We shall then base on that data to build an algorithm that predicts the suicide rates given the GDP per capita.



# **References.**

1. <https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa>
2. <https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook>
3. <https://www.eurekalert.org/pub_releases/2019-06/uot-sra062119.php>