

KABARAK UNIVERSITY

NAME:

GROUP 7

2. Benazir Salha

3.Becky Blessing

4.Dennis Kipkoech

5.Hamisi Salim

COURSE TITLE:

**ARTIFICIAL INTELLIGENCE AND
DATA SCIENCE**

**PROJECT TITLE: DATA
COLLECTION TOOLS AT NACSCOP**

DEDICATION

I dedicate this project to the health IT fraternity as well as the KABUO team that organized this training.

ACKNOWLEDGEMENTS

We thank the Almighty God for enabling me accomplish this task. We thank KABUO team for the advice and organization of this program . We also each other as a team for the joint efforts that got us this far.

DECLARATION

We hereby declare that this proposal report is based on our original work except for citations and quotations which have been duly acknowledged. We declare that the project hasn't been presented before.

Signature: _____

Name: Group leader(Samuel)

Date:_____

INTRODUCTION

1.1 INTRODUCTION

Nacscop is an organization that leads in response to HIV and STIs. It works in collaboration with the counties and therefore the hospitals within that county. Medical

institutions input their data initially using a tool known as ART which ran till 2018 when there was creation of a more detailed tool known as MOH 731. However, both of the tools are currently under use as transition is taking place gradually. The use of two systems is creating inefficiency and inaccuracy in data analysis by the organization.

1.2 PROBLEM STATEMENT

Data collection tools in NACSCOP are in transition. During the transition two sets are in use for collection and reporting. There's need to have an objective approach to select the indicator to use devoid of human bias and that includes other pertinent facility characteristics.

1.3 OBJECTIVES

1. To create a standardized, user friendly tool for data collection on HIV/AIDS and STIs in healthcare facilities countrywide.
2. To improve the quality of data collected on HIV and STIs by standardizing data collection processes and reducing errors in data entry.
3. To ease the data collection processes making it faster and highly efficient for healthcare workers.
4. To improve the accuracy and completeness of data on HIV/AIDS and STIs which can lead to better targeting and intervention and resources.
5. To improve the data management and analysis capabilities, enabling better decision making and

monitoring of progress towards national and global HIV/AIDS and STI goals.

1.4 JUSTIFICATION

The system solves the problem of the dual use of systems and acts as an effective decision-making tool. It accesses the situation a particular institution and chooses the best to for use.

1.5 SCHEDULE

This project has taken an estimated period of 3 days from 17th February to 20th February.

DAY 1: Being a day of accessing the problem, acquiring of data and establishing what possible solutions are there

DAY 2: Commencing of the project work; data cleaning, feature engineering and model development

DAY 3: Model evaluation/ testing and deployment.

CHAPTER TWO: METHODOLOGY

DESIGN METHODOLOGY

PROCESS DESIGN

Process design plays an important role in project development. In order to understand the working procedure, process design is necessary.

1. DATA EXPLORATION

The data obtained was studied in depth in order to gain insight on the trends. This is meant to inform the type of decision to be made and system to be developed.

2. DATA CLEANING

Our data cleaning process included:

- Removal of duplicates
- Replacement of null values with zero.
- Correcting of any existing errors

Cleaning out the data led to :

- Improved decision making .
- Better accuracy on results.
- Avoidance of repetitions.
- Saved on time.
- Minimization of any errors.

3. FEATURE ENGINEERING

- . Unpivoting the data into vertical data
- , Attribute selection
- .

4. MODEL DEVELOPMENT

- . Import all necessary libraries
- .Splitting the data into training and test sets
- . Print the size of the trainin and test sets
- .
- .
- .

: SYSTEM IMPLEMENTATION AND TESTING

6.1 INTRODUCTION

The project was a standalone application system and therefore was created using python programming language. This system depends on a database for storage of information .

The system was tested to find out whether it met the system requirements. It was meant to give an insight for the business entities to move from the traditional manual performance calculation system to an automated System. Development of a prototype was done to validate the system specification. The requirements were used to develop prototype software.

Tools used for testing and coding

- Google Colab – used as our coding environment.
-

LIMITATIONS, CONCLUSIONS AND RECOMMENDATION