



AFRICAN CENTER OF EXCELLENCE IN DATA SCIENCE



COLLEGE OF BUSINESS & ECONOMICS

**PREDICTION OF HIV INFECTIONS AMONG INDIVIDUALS WITH SEXUAL RISK BEHAVIOURS
IN RWANDA USING MACHINE LEARNING ALGORITHMS**

**By
Lorraine Muhimpundu
Registration number: 220000066**

**A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Data
Science in Biostatistics**


University of Rwanda, College of Business and Economics

**Supervisor: Dr. Pierre Claver Rutayisire
September 2022**

Declaration

I declare that this dissertation entitled “**Prediction of HIV infections among individuals with sexual risk behaviours using machine learning algorithms**” is the result of my work and has not been submitted for any other degree to the Centre for Data Science (UR-CBE) or any other university as a partial fulfilment for the degree of Master of Data Science in Biostatistics.

Name : Lorraine Muhimpundu

Signature : 

APPROVAL SHEET

This dissertation entitled '**Prediction of HIV infections among individuals with sexual risk behaviors using machine learning algorithms**' written and submitted by Lorraine Muhimpundu in partial fulfillment of the requirements of a degree for Master of Science in Data Science majoring in Biostatistics is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 19 % which is less than 20% accepted by the African Centre of Excellence in Data Science (ACE-DS).



Dr. Pierre Claver Rutayisire
Supervisor

Dr. Ignace H. KABANO
Head of Trainings

Dedication

This dissertation is dedicated to my family, parents, siblings, and friends who never ceased to encourage me throughout this journey. May God bless them all abundantly.

Acknowledgement

First and foremost, I would like to thank the Almighty God “JEHOVAH” for being with me and strengthening me the whole time. My gratitude goes to the Centre of Excellence of Data Science for granting me the opportunity to uptake this master’s degree in data science in Biostatistics and the lecturers who facilitated me through the whole process.

I am grateful for my supervisor, Dr. Pierre Claver Rutayisire who took his time to guide and keep me on the path throughout the dissertation writing journey regardless of his busy schedule. Special thanks to my family for their encouragement. Lastly, I appreciate my classmates’ great collaboration.

Abstract

The first case of HIV was identified back in 1920 in Congo and since then it has claimed over 32 million lives. Around 62% of new HIV infections occur among key populations and their sexual partners, including men who have sex with men (MSM), Female Sex Workers (FSWs), People Who Inject Drugs (PWID) and people in prison, despite them constituting a very small proportion of the general population (Data, 2019) and mostly because this population is made of all group of people that practice sexual risk behaviours which include inconsistent use of condoms, having multiple sexual partners, and paid sex in addition to early sex initiation. In Rwanda, HIV prevalence accounts for 3% among general population, 45.8% among female sex workers and 4.4% among men who have sex with men. This study aimed at building a model that on predicts new HIV infections among individuals with sexual risk behaviours by using the algorithms of machine learning. The study used 3 categories of variables (dependent or response variable, risk factors as independent variables, and demographic factors as independent variables as well). Data used were from the RPHIA dataset 2018-2019. Among 30,709 respondents, 29,775 (99.97%) were HIV negative and only 934 (0.03%) were HIV positive. Three machine learning classification algorithms namely logistic regression, gradient boost, and random tree forest were trained to find out the model that best predicts new HIV infections among individuals who practice sexual risk behaviours the random tree forest was found to be the best model with an accuracy of 71.15%, precision of 61.2%, recall of 84.5%, and F1-score of 70.9 at 0.35 threshold. obtained and predicted values were 261 true negatives, 163 false positives, 47 false negatives, and 257 true positives. Using random tree forest, it was observed that it minimizes the false negatives, increases true positives, recall and F1-score and the area under curve was 0.75. Feature importance was performed to determine the risk factors that influence new HIV infections occurrence among individual's who practice sexual risk behaviours and among social demographic variables, being in the age group of 15-24, being widowed or single, and having primary level of education were found to be factors that influence the HIV infection. While not having used condoms during last sexual intercourse, having debuted sex at an early age (under 20), and having multiple sexual partners (>1) were revealed to be risk behaviours that highly influenced the model that predicts HIV infections. This model will be essential for health public practitioners especially those who are most involved in HIV programs to design new programs about HIV prevention and transmission methods with emphasis on improving safe sex

negotiation skills and put more effort on educating young and adolescent children using the nationally approved ASRHR curriculum.

Keywords: Sexual Risk Behaviours, HIV, Key population, Machine learning

TABLE OF CONTENTS

Declaration	i
APPROVAL SHEET	ii
Dedication	iii
Acknowledgement	iv
Abstract	v
List of figures	ix
List of tables	x
Abbreviations	xi
Chapter 1 INTRODUCTION	1
1.1. BACKGROUND OF THE STUDY	1
1.2. Problem statement	2
1.3. Significance of the study	3
1.4. Research objectives	4
1.5. Research questions	4
1.6. Research hypotheses	4
Chapter 2 LITERATURE REVIEW AND THEORETICAL FRAMEWORK	5
2.1. Review of Literature	5
2.1.1. HIV/AIDS, sexual risk behaviours, MSM, FSWs, and PWID	5
2.1.2. Relationship between sexual risk behaviour and MSM, FSWs, and PWID ..	8
2.1.3. Existing machine learning HIV models related	9
2.2. Theoretical and Conceptual framework	11
2.2.1. Theoretical framework	11
2.2.2. Conceptual framework	12
Chapter 3 RESEARCH METHODOLOGY	14
3.1. Data source	14
3.2. Study population	14
3.3. Ethical considerations	14
3.4. Variables	15
3.5. Data Processing	18
3.6. Data Analysis	19
Chapter 4 DATA ANALYSIS	19
4.1. Descriptions of data	19
4.1.1. Demographic characteristics	19
4.1.2. Sexual risk behaviours	23
4.2. Distribution of socio-demographic characteristics and sexual risk behaviours by HIV status	

4.3. Predicting new HIV infections using machine learning algorithms	28
4.3.1. Logistic regression	28
4.3.2. RANDOM TREE FOREST	30
4.3.3. GRADIENT_BOOST	32
Chapter 5 DISCUSSION	33
5.1. Comparison of logistic regression, random forest, and Gradient Boost algorithms... ..	33
5.2. Social demographic-sexual risk behaviours-Random tree forest.....	34
5.3. Limitations of the study	36
Chapter 6 CONCLUSION	38
Chapter 7 RECOMMENDATIONS	39
REFERENCES	40
APPENDIX.....	42

List of figures

Figure 1. Modified Socio Ecological Model.....	11
Figure 2. Conceptual framework	13
Figure 3. HIV status distribution.....	19
Figure 4. Age distribution	20
Figure 5. Gender distribution	20
Figure 6. Marital status distribution.....	20
Figure 7. Education level distribution.....	21
Figure 8. Employment status distribution	21
Figure 9. Wealth quintile distribution	22
Figure 10. Residence type distribution.....	22
Figure 11. Age at first sex distribution.....	23
Figure 12. Male circumcision status distribution.....	23
Figure 13. Number of sexual partners distribution	24
Figure 14. Condom use distribution.....	24
Figure 15. Relationship with sexual partners distribution	24
Figure 16. Anal sex distribution.....	25
Figure 17. Logistic regression-Confusion matrix at 0.5 threshold.....	29
Figure 18. Logistic regression-Confusion matrix at 0.35 threshold.....	29
Figure 19. Logistic regression-ROC curve at 0.35 threshold.....	30
Figure 20. Random tree forest-Confusion matrix at 0.5 threshold.	30
Figure 21. Random tree forest-Confusion matrix at 0.35 threshold.	31
Figure 22. Random tree forest-Roc curve at 0.35 threshold	31
Figure 23. Gradient boost-Confusion matrix at 0.35 threshold.	32
Figure 24. Gradient boost-Roc curve at 0.35 threshold	32
Figure 25. Feature importance techniques	49

Abbreviations

ACE:	African Center of Excellence
AIDS:	Acquired Immuno-Deficiency Syndrome
BMC:	Bio-Medical Central
CBE:	College of Business and Economics
CD4:	cluster of differentiation 4
CI:	Confidence Interval
DHS:	Demographic Health Survey
DS:	Data Science
EA:	East Africa
FSW:	Female sex Worker
GoR:	Government of Rwanda
HH:	Household
HIV:	Human Immunodeficiency Virus
ICAP:	International Centre for AIDS care and treatment Program
MSEM:	Modified Socio Ecological Model
MSM:	Men who have sex with men
NISR:	National Institute Statistics of Rwanda
PWID:	People Who Inject Drugs
RBC:	Rwanda Biomedical Centre
RDHS:	Rwanda Demographic Health survey
ROC:	Receiver Operating Curve
RPHIA:	Rwanda Population based HIV Impact Assessment
SRHR:	Sexual Reproductive Health and Rights
UNAIDS:	United Nations Programme on HIV/AIDS
UR:	University of Rwanda
USA:	United States of America
WHO:	World Health Organization

List of tables

Table 1: List and type of variables15
Table 2: Distribution of socio-demographic characteristics and sexual risk behaviours
by HIV status27

Chapter 1 INTRODUCTION

1.1. BACKGROUND OF THE STUDY

HIV remains a major global public health issue that claimed over 32 million of lives so far. According to Global HIV statistics (avert.org); worldwide, 38 million of people are infected by HIV, whereby 68% of them live in the sub-Saharan Africa and 26 million of them live in east and southern Africa. (Data, 2019)

In its report Global AIDS Epidemic 2020 report, UNAIDS revealed that the majority of new HIV infections occurred among key populations and their sexual partners, including men who have sex with men (MSM), Female Sex Workers (FSWs), People Who Inject Drugs (PWID) and people in prison, despite them constituting a very small proportion of the general population, they represented 62% (Data, 2019). After being diagnosed HIV positive, these people do not practice safer sex even though they are fully aware of their seropositivity status. According to Fisher JD, Smith L., Secondary prevention of *HIV* infection: the current state of prevention for positives unprotected sex continues to account for the largest proportion (80%) of new HIV infections globally (Jeffrey D. Fisher & Smith, 2010).

As it is defined in DHS 2015, sexual risk behaviours are behaviours practiced by a certain group of people which might put them on risk of getting infected by HIV or behaviours people with HIV negative known status that put them at high risk of getting infected by HIV. The risk behaviours include inconsistent use of condoms, having multiple partners, early initiation of sex, and paid sex. (Rwanda, 2015)

In Rwanda, while the HIV prevalence remains constantly 3%, the prevalence among female sex workers is 45.8% and 4.4% among gay men (Rwanda, 2015). Regarding HIV prevalence disaggregated by risk behaviours, DHS 2015 revealed that HIV prevalence is high among men whose sexual debut was at age 18-19 (4 percent) and lowest among those whose sexual debut was before age 16 (less than 1 percent). According to Rwanda (2015), the prevalence of HIV infections among women who have had one romantic partner rose from 3 to 13 percent for those who have had three to four romantic partners, whereas it varied from 1 percent for men who have had one romantic partner to 13 percent for men who have had 10 or more partners. Moreover, it varies from 1% among men who have only ever had one lifetime partner to 13% among those who have had at least ten. Those who reported paying for sex or engaging

in sexual activity within the previous 12 months have a slightly higher rate of HIV infection than men who did not (5 percent versus 3 percent). Participants in the survey who have had more than one sexual partner have a greater prevalence of HIV (6% of respondents with two or more partners and 8% of those with concurrent partners in the previous year are HIV-positive). (Rwanda, 2015).

1.2. Problem statement

The primary cause of the HIV epidemic in South Africa is sexual behavior (Hallman, 2005). Between the ages of 12 and 14, young South Africans start acting sexually, and before the age of 16, 11% of males and 6% of females make their first sexual intercourse (Mathews, 2010). Men who were sexually active and aged 15 to 24 (31%) reported having numerous partners in the preceding 12 months. Condoms were not used in up to 40% of the most recent instances of sexual contact among sexually active women and men. 2010 (Mathews).

The main cause of the extended spread of HIV, according to Crepaz N. and Marks G., is people who are aware of their HIV seropositivity but continue to engage in risky sexual behaviors. Furthermore, many HIV-positive people engage in safer sex practices, despite the fact that a sizeable proportion of seropositive people still participate in risky sexual behaviors that expose others to infection. (Crepaz N, 202).

According to DHS 2015 findings, there is a high HIV prevalence in Rwanda among men who had their first sexual experience between the ages of 18 and 19 (4%) and men and women who have multiple sexual partners (6%) as well as among men who admitted to having paid for a sexual encounter within the previous 12 months of the survey. (Rwanda, 2015).

Although so many HIV prevention programs have been implemented, the prevalence of HIV remains unchanged, and this is mostly due to high prevalence of HIV found among the group of people who are more likely to practice sexual risk behaviours.

In Rwanda, so many studies have been conducted on HIV prevention among the priority population known as FSWs, MSM and PWID who have sexual risk behaviours, however, our current knowledge, there has not been any study in Rwanda about the

prediction of new HIV infections among these groups according to the risk behaviours practiced.

It is in this regard that there is a need to conduct the study on prediction of new HIV infections among individuals with sexual risk behaviours by using the algorithms of machine learning.

1.3. Significance of the study

In Rwanda, HIV Prevalence remains unchanged which is most likely due to the elevated prevalence of HIV among individuals with sexual risk behaviors who are mainly men who have sex with men with 4%, people with multiple sexual partners known as female sex workers (and their clients) with 45.8% (Rwanda, 2015), and people who inject drugs who are 22 times more at risk of HIV compared to the general population (Data, 2019). Besides, various studies on HIV were conducted, however, no similar studies on prediction of new HIV infections among individuals with sexual risk behaviors using machine learning algorithm was performed.

By using the machine learning algorithms, a model that predicts the new HIV infections among population with sexual risk behaviours will be built and will be used to test data for different period. Machine learning techniques are more preferable to use when dealing with big dataset than traditional statistical methods. Machine Learning was chosen since according to Rajula, Manchia, Giuseppe, & Antonucchi in their journal on comparison of conventional statistics methods with machine learning in medicine, it focuses on making predictions as accurate as possible, while traditional statistical models intend at concluding relationships between variables. Machine learning conveniences include flexibility and scalability compared with the methods of conventional statistical that make it deployable for several tasks such as diagnosis and classification, and survival predictions.

The study results will reveal the potential sexual risk factors associated with occurrence of new HIV infections and a predictive model will be built using machine learning. Policy makers, decision makers and other public health practitioners will use the findings to develop new strategies and design new interventions in line with HIV prevention programs and in making decisions about future HIV prevention programs. The predictive model will be used in future to predict new HIV infections.

1.4. Research objectives

The overall objective of this study is to build a model that predicts the occurrence of new HIV infections and people with sexual risk behaviours.

Below are the specific objectives:

- ✓ Identify potential risk factors that influence the model that predicts HIV infections among sexual risk behaviours individuals.
- ✓ Develop a random forest model that will predict new HIV infections among individuals with sexual risk behaviours.

1.5. Research questions

Below are the questions that the research will respond to:

1. What are the risk factors that influence the prediction of HIV infections among sexual risk behaviour individuals?
2. Does the model predict accurately the occurrence of new HIV infections among individuals who engage in sexually risky behaviors??

1.6. Research hypotheses

a) Null hypothesis:

- there are no risk factors that influence the model that predicts HIV infections and sexual risk behaviours.
- There is no model that accurately predicts HIV infection among people with sexual risk behaviours.

b) Alternative hypothesis:

- Sexual risk behaviours factors influence the model that predicts HIV infection.
- There is a model that accurately predicts HIV infection among people with sexual risk behaviours

Chapter 2 LITERATURE REVIEW AND THEORETICAL FRAMEWORK

This chapter entails a relevant literature review to this study. The first section will provide a review on HIV/AIDS, sexual risk behaviours, people who are likely to practice sexual risk behaviours, the relationship between people with sexual risk behaviours and sexual risk behaviours practiced, and finally pertinent HIV models and machine learning methods used will be discussed. The second session contains a theoretical and conceptual framework which will help us to understand the variables involved in the study being conducted.

2.1. Review of Literature

2.1.1. HIV/AIDS, sexual risk behaviours, MSM, FSWs, and PWID

HIV/AIDS

The 1st case of HIV was found in Congo in 1920s according to Avert journal in its article origin of HIV/AIDS, but people became aware of it in 1980s in USA when few cases of rare diseases (Kaposi's Sarcoma (a rare cancer) and a lung infection were reported among gay men, and it was concluded that there was an infectious disease that was causing them. That infectious disease was then confirmed to be a new health condition. (Avert, 2017).

There are four main ways in which HIV is transmitted; having unprotected sexual intercourse (vaginal or anal) with an HIV positive person, sharing of unsterilized sharp objects such as injecting drugs equipment, transfusion of infected blood, and mother to child transmission that can happen during pregnancy, childbirth, or breastfeeding.

HIV/AIDS continues to be an epidemic with an estimate of 38 million of people who live with HIV and 32 million people who have already died which results in 70 million HIV infections as of 2019.

Globally, key population known as gay men, those involved in paid sex and their clients, people who inject drugs, and transgender people, accounted for 62% of HIV new infections with 23%, 19%, 10%, and 8% for men who have sex with men, clients of sex workers and their partners, people who inject drugs, and sex workers respectively (UNAIDS, 2020). This population is made of all group of people that practice sexual risk behaviours which include inconsistent use of condoms, having multiple sexual

partners, and paid sex in addition to early sex initiation. (DHS Rwanda, 2015). Although the mentioned sexual risk behaviours are claimed to be more practiced by key population, general population also practice same behaviours the reason why HIV/AIDS is present among general population as well.

Sexual risk behaviours

Sexual risk behaviours were defined by Leventhal as behaviours that increase susceptibility of an individual to problems related to sexuality and reproductive health (Leventhal AM, 2017). Early initiation sex, having multiple sexual partners, having sex while under the influence of alcohol or drugs and unprotected sex are the common characteristics of risky sexual behaviours which increases risk of individuals to sexuality and reproductive health problems (Kebede, Bogale, & Gerensea, 2017). In addition, Alem Girmay* and Teklewoini Mariye in their study “Risky sexual behaviour practice and associated factors among secondary and preparatory school students of Aksum town, northern Ethiopia, 2018”, defined sexual risk behaviours among students as when a student engage in one of the behaviours which are multiple sexual partners (having more than one sexual partner until the survey), start engaging in sex at the age < 18 years old, inconsistent use of condom (inconsistent/fail to use condom at least one during sexual intercourse until the survey), and having sex with commercial sex workers at least once until the survey (Alem & Teklewoini, 2018). The above definition of sexual risky behaviours applies to both general population and key population especially that students are part of the general population.

Furthermore, sexual risk behaviours are not practiced by only HIV negative individuals, HIV positive persons also engage in those behaviours. In Ghana, a study on sexual risk behavior among HIV-positive people found that 44 percent of the participants had sex after testing positive for the virus, 67 percent were in relationships with unknowing or negative partners, and more than half (51 percent) of the study population with regular sex partners said they engaged in unprotected anal or vaginal sex. (Jolly D. P., 2012).

Men who have sex with men

Men who have sex with men (MSM) is a term used when a man is aroused by another man.

In line with capturing by sexual orientation (homosexual, bisexual, heterosexual, or

gay) and gender identity (male, female, transgender, queer) a range of male-male sexual behaviors and avoiding characterization of men practicing these behaviours, the term MSM was established, and this was in 1992. (Prof. Chris Beyrer, et al., 2012). The term Briefly, the term is used to categorize men that participate in sexual activity with other men regardless of their sexual orientation or identification. Men who identify as gay, heterosexually identified men who have sex with men, bisexual men, male sex workers of any orientation, men who engage in these behaviors in all settings where men are present, such as prisons, and the rich and diverse range of traditional identities and terms for these men across cultures and subcultures make up the MSM community. (Prof. Chris Beyrer, et al., 2012).

Female sex workers

Cheryl Overs. 2002, Sex workers are women, men, and transgender people who get money or items in return for sexual services and who consciously characterize those acts as revenue producing even though they do not consider sex work to be their employment (Overs,2002). In this study we will consider female sex workers including girls' adolescents who are sexually active.

Avert journal in its article sex workers revealed that on average, sex workers are 13 times more likely to become infected with HIV than adults in the general population and that sex workers make up 9% of the total number of new HIV infections across the world. (Sex workers, HIV and AIDS, 2019)

People who inject drugs

People who inject drugs inject narcotics into their bodies using syringes and needles. World Health Organization ((WHO), 2015) in its technical brief defines people who inject drugs as people who inject non-medically sanctioned psychotropic (or psychoactive) substances. These drugs include, but are not limited to, opioids, amphetamine-type stimulants, cocaine, hypno-sedatives and hallucinogens (2015, p. 38). Injection may be through intravenous, intramuscular, subcutaneous, or other injectable routes ((WHO), 2015). In Rwanda, the most used injectable drugs are heroine, cocaine, and ketamine. Population of people who inject drugs include FSWs, MSM, and people from the general population.

People who inject drugs are considered as a population which is at higher risk of getting infected with HIV than the general population (Ochonye, et al., 2019). This is because these individuals share needles and once a person who is HIV positive uses the needle, he/she put others who are the HIV negative at higher risk of getting infected and sharing injectables needles and sharp objects is one of the main ways in which HIV is propagated.

2.1.2. Relationship between sexual risk behaviour and MSM, FSWs, and PWID

According to UNAIDS epidemiological estimates, 2019; the proportion of new adult HIV infections among key populations and their sexual partners worldwide was 62%, and the relative risk of HIV infection acquisition compared to the rest of the population is 13 times higher for transgender people, 26 times higher for gay men, 29 times higher for people who inject drugs, and 30 times higher for female sex workers.

According to the findings from behavioral and biological surveillance survey among female sex workers conducted by TRAC Plus; overall, only 33% of FSW reported having consistently used condoms with the paying sexual partner in the last month preceding the survey and that the majority of FSW (65%) reported having three or more paying partners in the last seven days FSWs. Because of the nature of their work, FSWs are exposed to substance use (alcohol and other drugs) for them to be able to engage in sexual activities especially that most of them involve in paid sexual activities against their willingness rather because of poverty. A study in China showed that 29.4% of the sampled FSWs used alcohol before sex. Once they use alcohol or any other drug, they lose control and condom use negotiation which put them at high risk of HIV infection. In addition, there are men who prefer not using condoms and offer a huge amount of money that the sex work cannot resist. one-third of FSWs reported having had sex without a condom because clients paid more money or looked clean (Cai, et al., 2010)

Moreover, (Ochonye, et al., 2019) mentioned in his study that FSWs, MSM, and PWID are more likely to engage in sexual risk behaviour, findings more revealed that 71.9% of the population had their first sexual intercourse when they were still adolescents (10–19 years). The percentage of people that use psychoactive drugs and inject drugs was high as well in MSM and FSWs. Injection of drugs was found to be the highest HIV risk behaviour since PWID lower chance of using condom during sexual intercourse and refuse sex with partners who refused to use condoms when

compared with FSW and MSM; and more than a third of the population share needles and syringes. A study on risk prediction score for HIV infection: Development and internal validation with cross-sectional data from men who have sex with men conducted in China revealed that multiple sexual partners, not using condom consistently, alcohol use, and aphrodisiac substances were associated with increasing HIV transmission risk among Chinese MSM (Yin, et al., 2018).

Besides, injecting drugs and substance use/alcohol use often goes with increased sexual desire, decision making inability, and lowered self-consciousness behaviours which can influence having unprotected sex and once it becomes a habit, you may end up having unprotected sexual intercourse with multiple partners. A technical brief (HIV and Young People Who Inject Drugs) claims that young people who inject drugs are more susceptible to HIV infection and have a variety of other vulnerabilities. They may perform sex acts in exchange for drugs or gain funds to buy drugs, making it challenging for them to decline a transaction or insist that a customer wear a condom and raising their risk of catching HIV. Their vulnerability to abuse, including rape, was also noted in the technical brief. ((WHO), 2015).

In Ghana, a considerable percentage of HIV positive individuals engage in unprotected sexual intercourse and are most likely to put others at risk of HIV infection. More than half (51%) of the study group with regular sex partners reported having unprotected anal or vaginal sex. Participants who self-reported not using condoms during their most recent sexual encounter were 31 percent. (N.M.NCUBE, et al., 2012).

2.1.3. Existing machine learning HIV models related

In Rwanda, various surveys have been conducted to find the prevalence of HIV in both general population and key populations as well as nationwide, and to estimate the annual incidence of HIV among adults (15-64 years old). The surveys include Rwanda Demographic Health Surveys (RDHS), and Rwanda Population HIV Impact Assessment (RPHIA). The DHS reports revealed that key population (MSM & FSWs) are more likely to engage in sexual risk behaviours and that general population also engage in risky behaviours that can put them at risk of HIV infections; However, no study that predicts new HIV infections among individuals who practice sexual risk behaviours was conducted.

A study by Tary Nicole Ho Tim on predicting HIV status using neural network and demographic factors, obtained a classifier of HIV status of a patient based of demographic data by using neural network to perform knowledge discovery and data mining on HIV clinical and demographic data. The study uncovered that the average accuracy was between 61% and 62% by using neural network trained with Bayesian classifier evidencing that demographic factors such as race, region, age of the mother, age of the father, education level of the mother, gravidity, parity, and HIV status; are not accurate predictors of HIV status. (Tim, 2006)

In addition, A study by Yashik Singh¹, Nitesh Narsai, and Maurice Mars titled "Applying machine learning to predict patient specific current CD4 cell count in order to determine the progression of human immunodeficiency virus (HIV) infection" also found that CD4 cell count measurements can be accurately predicted using machine learning. Using protease and reverse transcriptase genomes, viral load, and the length of weeks since baseline measurement, it is feasible to forecast actual CD4 cell counts and determine if a patient's CD4 cell count is fewer than 200. Support vector machines were shown to outperform neural networks, both in predicting CD4 cell count and if the CD4 cell counts are less than 200. (Singh, Narsai, & Mars, 2013).

A study by Erol Orel "Machine learning to identify socio-behavioural predictors of HIV positivity in East and Southern Africa" found that XG Boost algorithm performs best the prediction of HIV infections with a mean score of 76.8% [95% CI 76.0%-77.6%] for males and 78.8% [78.2%-79.4%] for females. In addition, a direction of the relationship between HIV status and its forecasters was determined and results showed that older age of the recent partners, large number of sexual partners, living in urban district were found to be HIV risk factors and having had previously used condom increased HIV positivity; while, circumcision, and breastfeeding were associated with lower risk of HIV infection. (Orel, et al., 2020)

Moreover, research on use of Machine Learning Techniques to Identify HIV Predictors for Screening found out that using XGBoost; age, avoiding pregnancy, living in urban districts, low level of education, TB, and being uncircumcised were the behaviours more related to HIV positivity. (McSharry, Mutai, Ngaruye, & Musabanganji, 2021).

2.2. Theoretical and Conceptual framework

2.2.1. Theoretical framework

This study will use a part of modified social ecological model (MSEM) expanded by Stefan Baral, Carmen H Logie, Ashley Grosso, Andrea L Wirtz, and Chris Beyrer (2013). This model was built on previous frameworks such as health belief model which focuses on understanding and predicting detection of diseases, theory of planned action that explains all behaviours that people might engage in and are able to use self-control, the model of behaviour changes (that explains why human behaviours change), and social ecological model which explains the complex associations between social and structural factors, individual practices, and the physical environment and health. (Baral, Logie, Grosso, Wirtz, & Beyrer, 2013)

The MSEM of HIV epidemic model is made of 5 levels of HIV infection risks which are: 1) individual, 2) network, 3) community, 4) policy, and 5) stage of HIV epidemic. MSEM is a modification of social ecological model which was changed by adding the fifth layer of HIV epidemic based on the assumption that though individual risk behaviours are essential to spread a disease but are not enough to explain dynamics of the epidemic at population level. (Baral, Logie, Grosso, Wirtz, & Beyrer, 2013)

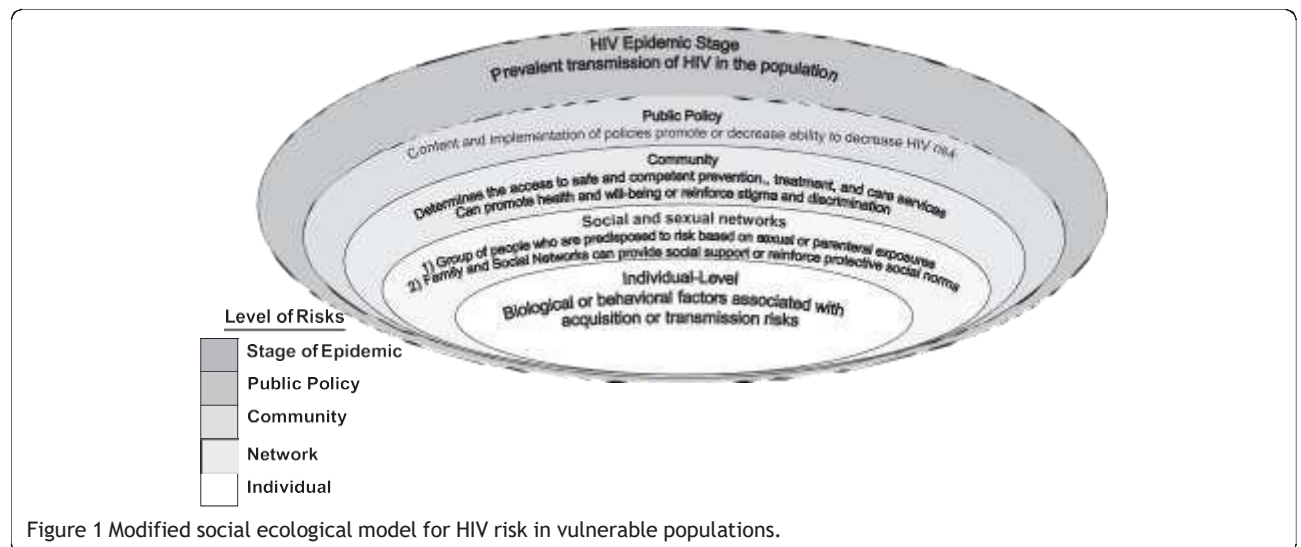


Figure 1. Modified Socio Ecological Model

Individual factors are biological or individual factors that influence the acquisition of a disease as well as its transmission. HIV risk individual factors include age, condom use, substance use, multiple sexual partners.

Networks consist of liaison or interrelationship that a person has for instance, family, friends, or neighbors that can directly influence behaviours that put him/her at risk of acquiring a disease, for example HIV infection in our study. (Baral, Logie, Grosso, Wirtz, & Beyrer, 2013).

Community is a place in which live people who are bound by geographical, socioeconomic, socio-economic status, cultural or racial and religion (Baral, Logie, Grosso, Wirtz, & Beyrer, 2013) In a community, a person may be influenced by the community's behaviours that can expose him to HIV infections.

Public policy is crucial in slowing HIV infection and supporting HIV prevention programs among marginalized and the general population by providing funds to legal policies which are implemented to promote harm reduction services (e.g.: condom provision, needle exchange) within the community by making those actions illegal or legal depending on the nature of the action.

Stage of HIV epidemic is derived from individual characteristics, socio-networks behaviours, community influence behaviours, and policy that can define the risk of getting HIV at individual level. However, according to Baral et al. 2013, no behaviour, policy or law, community determinant, network attribute, or individual characteristic can create infectious disease; rather these can only create conditions which either increase or decrease the probability of acquisition or onward transmission of an already prevalent disease.

2.2.2. Conceptual framework

This study will use 3 parts of the MSEM model which are: individual factors, social networks, and community.

Individual factors as mentioned above are individual behaviours that a person can use consciously that put them at risk of acquiring HIV. The individual behaviours include sex age debut, inconsistent use of condoms, number of sexual partners, unprotected sexual intercourse, substance use, age, marital status, and education level.

Social networks are comprised of interpersonal relationships which include family, friends, neighbours, and other networks that directly influence health behaviours (Baral, Logie, Grosso, Wirtz, & Beyrer, 2013). Since they are more prone to engage in sexually risky behaviors, FSWs, MSM, and drug users (PWID) are referred to as social networks in our study. However, the overall population with dangerous HIV behaviors will be taken into account

At community level, the study will consider the type of residence. Although, there are many community factors that may influence HIV infection, for instance stigma and discrimination, wealth status, region, and ethnicity.

The below chart shows that the more FSWs, MSM and PWID practice the sexual risk behaviours (inconsistent use of condoms, having multiple sexual partners, and paid sex intercourse), the more HIV infections will arise. The following framework demonstrates the causal effect relationship between the new HIV infections and sexual risk behaviours.

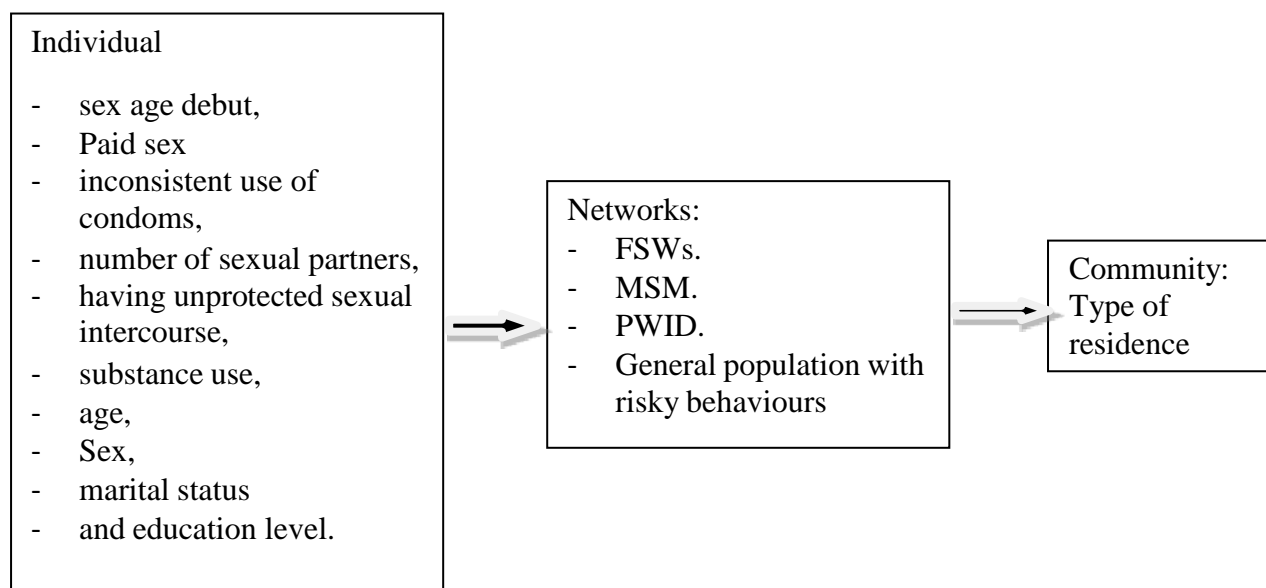


Figure 2. Conceptual framework

Chapter 3 RESEARCH METHODOLOGY

This chapter provides an overview of the methodology that will be used in this study. The chapter also summaries and talks about the data source, ethical consideration, population size selection, data processing, variable description, and the method for data analysis.

3.1. Data source

The study will make use of secondary data from the RPHIA survey, which was carried out by the International Center for AIDS Care and Treatment Program (ICAP) between October 2018 and March 2019 in collaboration with the Rwanda Biomedical Centre (RBC), the Ministry of Health (MoH), and the National Institute of Statistics (NISR).

3.2. Study population

A two-stage, stratified cluster sampling strategy was used to choose the dataset's 35,265 individuals. Based on the 2012 census, which included 10,378,021 people from 2,424,898 HHs throughout 16,728 enumeration zones, the target population was all individuals from all households (HHs) in Rwanda. The average number of HHs and persons per EA was 145 and 623 respectively.

During the first stage, a probability proportional to size method was used to select 375 EAs which were stratified by provinces. For the second stage, a random sampling with equal probability method was used to sample of HHs within each EA as well. On average, 30 HHs were sampled and the number of sampled HHs ranged between 14 and 60. All persons from the selected HHs were interviewed.

3.3. Ethical considerations

RPHIA is a national household-based survey conducted basing on the total population counted during Rwanda Demographic Household Survey. Data were accessed from ICAP online datasets where I had to first apply to get access to data. During access to data request, I described my research topic, provided the study relevancy, and explained how the data will be used. Access was granted via email and downloaded the data. The obtained data do not reveal individual's identity and they will solely be used for academic purposes.

3.4. Variables

This study will use several variables from the RPHIA dataset 2018-2019, and they are comprised of 3 groups: the dependent variable (response) and independent variables (risk factors, and demographic factors). Demographic factors include; type of residence, education level, marital status, gender, age, wealth quintile, risk factors include are you circumcised, age at first sex, number of sexual partners within last 12 months prior the survey, relationship with sexual partner, gender of the sexual partner, condom use at last sex, paid sex (with expectations: money, gifts, others)), anal sex, have you ever tested for HIV, and when were you tested HIV positive for the first time, HIV status, and finally the response variable is new infections.

Table 1: List and type of variables

Variable	Type of variable	Variable description	Coding
HIV status (243, 246)	Dependent	HIV negative HIV positive	1 –HIV Positive 2 - HIV Negative 99-Missing
Type of residence (241)	Independent	Does the respondent live in urban or rural area/region?	1=Urban 2=Rural
Education level (247,31)	Independent	What is the Respondents' level of education?	1 - No education 2 - Primary 3 - Secondary 4 - More than Secondary 99 - Missing
Marital status (37,250)	Independent	Current marital status: married, living together with someone as if married, widowed, divorced, or separated	1- married 2- living together 3 - widowed 4- divorced 5- separated 8- don't know 9 - refused
Gender (7)	Independent	Gender of the respondent: Male? Female?	1 - Male 2 - Female 99 - Missing
Age (8)	Independent	Integer	Integer (15... ..)

Wealth quintile (244)	Independent	Economic level of the respondents	1 - Lowest 2 - Second 3 - Middle 4 - Fourth 5 - Highest 99 - Missing
Are you circumcised? (105)	Independent	There are men who are uncomfortable talking about circumcision, but it is an important information for us to have. Some men are circumcised. Are you circumcised?	1 - Yes 2 - No -8 - Don't Know -9 - Refused
Age at first sex (109)	Independent	How old were you when you had vaginal sex for the very first time? Vaginal sex is when a penis enters a vagina	Integer (1... ..)
Age at first sex (110)		Please provide the reason this previous question was left blank: How old were you when you had vaginal sex for the very first time?	96 - Never Had Vaginal Sex -7 - Out of Range -8 - Don't Know -9 - Refused
Sexual partners (111)	Independent	Number of sexual partners (people they had sexual intercourse with) within the last 12 months. Note: The total number of sexual partners reported in the last 12 months may exceed the total number of reported lifetime sexual	Integer
Sexual partners (112)	Independent	Please provide the reason this previous question was	-8 – Don't Know -9 – Refused

		left blank: In total, how many different people have you had sex with in the previous 12 months?	
Relationship with sexual partners (114)	Independent	What is your relationship with your sexual partner?	1 – Husband or Wife 2 - Live-In Partner 3 - Partner, Not Living with Respondent 4 - Ex-Spouse/Ex-Partner 5 - Friend/Acquaintance 6 - Sex Worker 7 - Sex Worker Client 8 - Stranger 96 - Other (Specify) -8 - Don't Know -9 - Refused
Gender of the sexual partner (115)	Independent	Is your sexual partner male or female?	1 - Male 2 - Female -8 - Don't Know -9 - Refused
Condom use (118,273,274)	Independent	The last time you had sex with your sexual partner was a condom used?	1 - Yes 2 - No -8 - Don't Know -9 - Refused
Paid sex (in kind, money) (119)	Independent	Did you have sex with your sexual partner because he provided you with or you expected that [PARTNER'S NAME] would provide you gifts, help you to pay for things, or help you in other ways?	1 - Yes 2 - No -8 - Don't Know -9 - Refused
Anal sex (162)	Independent	There are different ways in which people have sex: vaginal sex and anal sex. Anal sex is when a penis enters a	1 - Yes 2 - No -8 - Don't Know -9 - Refused

		person's anus. Have you ever had anal sex?	
Ever tested for HIV? (179)	Independent	YEAR (Which month and year was your last HIV test?)	Year -8 - Don't know year -9 - Refused year
What was the result (181)	Independent	What was the result?	1 - Positive 2 - Negative
When were you tested HIV +for the first time (183)	Independent	YEAR (What was the month and year of your first HIV positive test result?)	Year -8 - Don't know year -9 - Refused year

a. Data Processing

The dataset was imbalanced whereby 97% of the cases were HIV negative and the ratio of HIV positives was 3%. This was a very big challenge since machine learning algorithms are sensitive to highly unequal classes and do not perform well with datasets of class imbalances. To reduce the size of the majority class, down sampling method was used by randomly selecting 4.5% of HIV negative people to ensure representativeness. This was done because when one class is underrepresented, machine learning algorithms do not get all required information for the minority class to make accurate information, and this may lead to biased predictions and accuracies in favour of the majority class.

Missing values were handled using 2 methods: 1) was to judge based on the previous or questions following the non-answered questions, for instance we had cases where a person who never had sex and never married reported that he/she didn't use condoms during last sexual intercourse and that he/she had sex with a partner or a friend, we imputed that he never had sex. therefore, all individual with the same case were imputed that they never had sex. 2) Another imputation method was to complete missing values using k-Nearest Neighbors whereby each sample's missing values were imputed using the mean value from n_neighbours nearest neighbors found in the training set. Two

samples are close if the features that neither is missing are close. Lastly, since the dataset had categorical data and machine learning to perform requires that all input and output variables are numeric, they were encoded to numeric values before evaluating and fitting the model. They were encoded using one-Hot encoding.

b. Data Analysis

Descriptive statistics of the demographic characteristics such as age, gender, marital status, region, and education level will be computed.

In addition, the study will also apply supervised learning algorithms such as linear regression, random tree forest, and gradient boost for classification by using python to build and to compare the accuracy of different models in order to come up with the model that best predicts new HIV infections, and the risky factors that influence in determining HIV seropositivity.

Chapter 4 DATA ANALYSIS

4.1. Descriptions of data

4.1.1. Demographic characteristics

HIV status

HIV prevalence among the surveyed participants 3% (934), while those who are HIV negative present 97% (29,775) of the sample population.

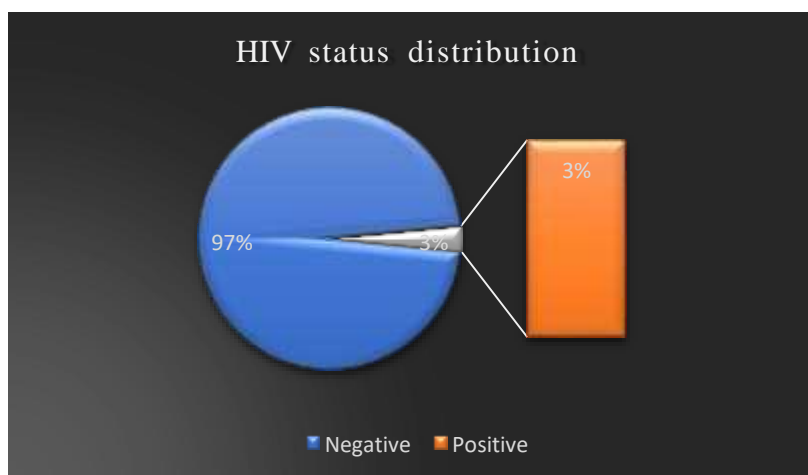


Figure 3. HIV status distribution

Age

Most of the surveyed population are aged between 15 and 24 years old who accounted for 37% followed by 25-34, 35-44, 45-54, and 55-64 that accounted for 27%, 18%, 10%, and 8% respectively.

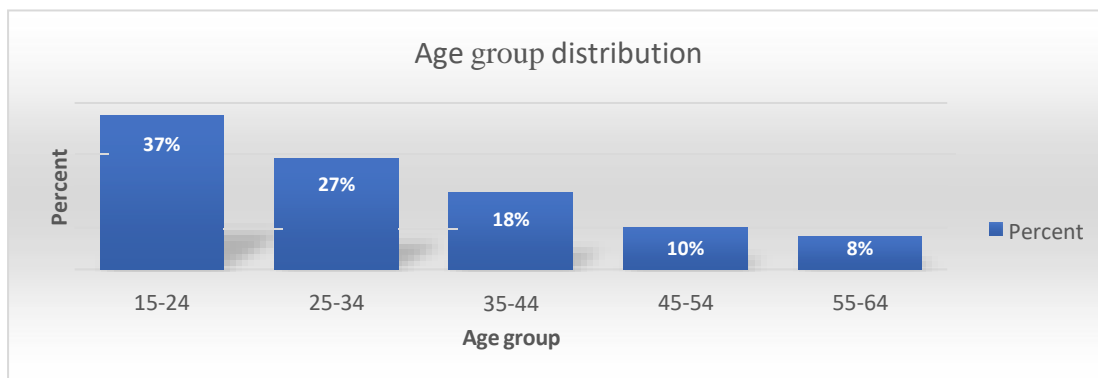
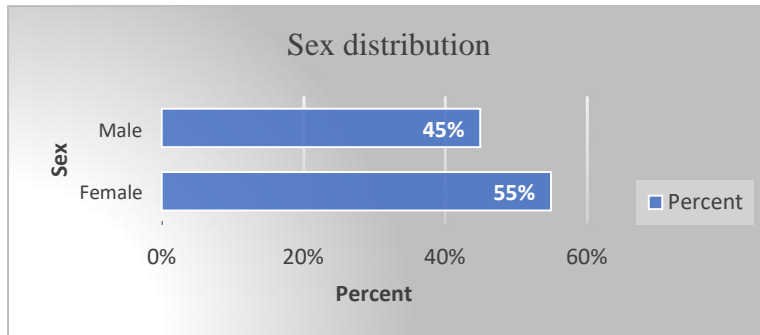


Figure 4. Age distribution

Gender

Most of beneficiaries are females (55%) and males were 45%



Marital status

Regarding marital status, the majority are married (48%) followed by those who never married (42%), and the least reported marital status was divorced which accounted for 1%. Separated and widowed accounted for 5% and 4% respectively.

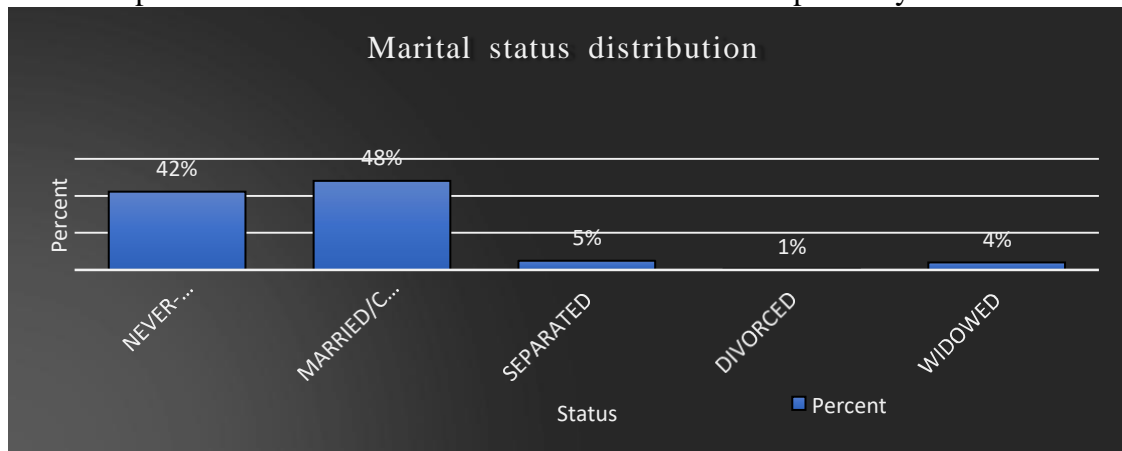
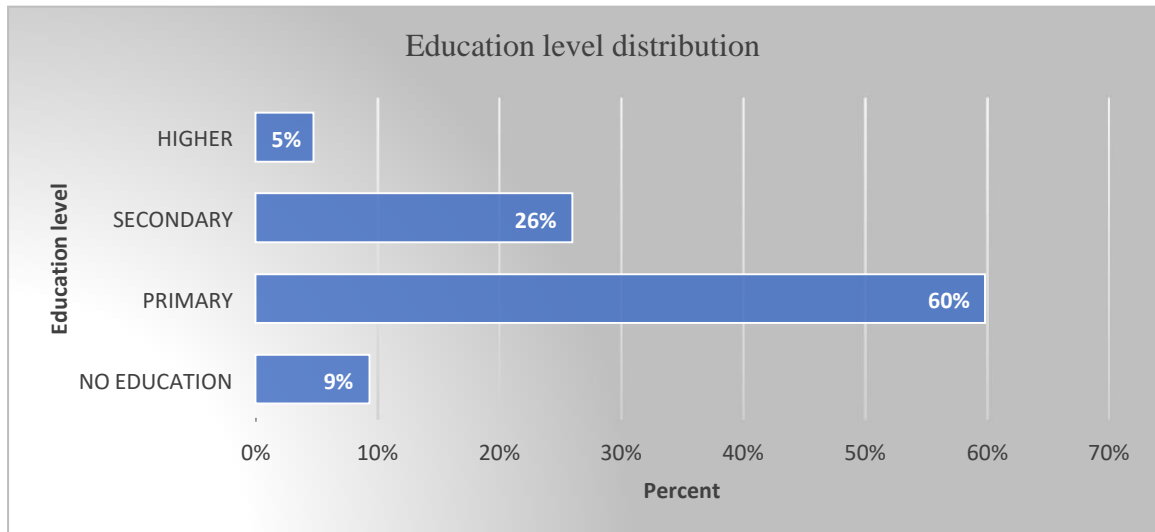


Figure 6. Marital status distribution

Education level

More than half of respondents had primary education (60%), those with secondary education were 26%, the least reported level of education was higher education that accounted for 5%. 9% of the sample population had no education.



Employment status

The below Venn diagram shows that most surveyed population were unemployed (60%) within the last 12 months prior survey, and those who are employed were 40%.



Figure 8. Employment status distribution

Wealth quintile

An increasing pattern was observed for wealth quintile distribution from lowest to fifth quintile with 17%, 18%, 19%, 20%, and 26% respectively.

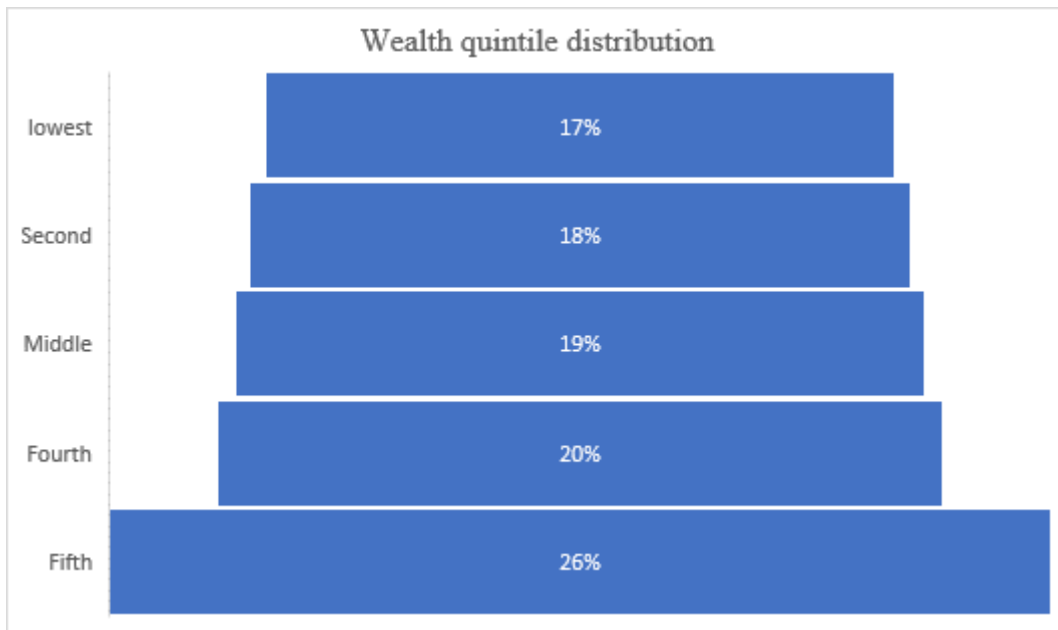
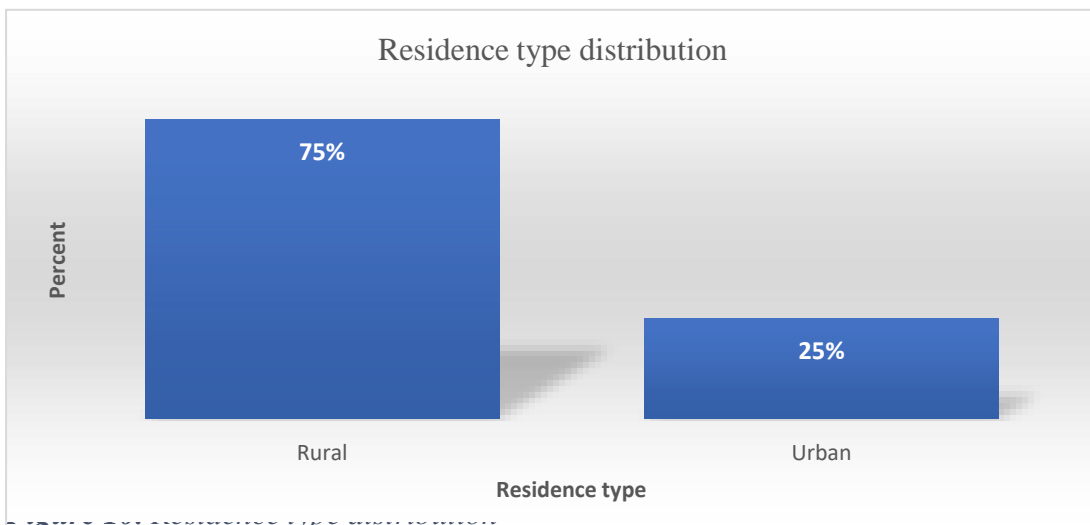


Figure 9. Wealth quintile distribution

Type of residence

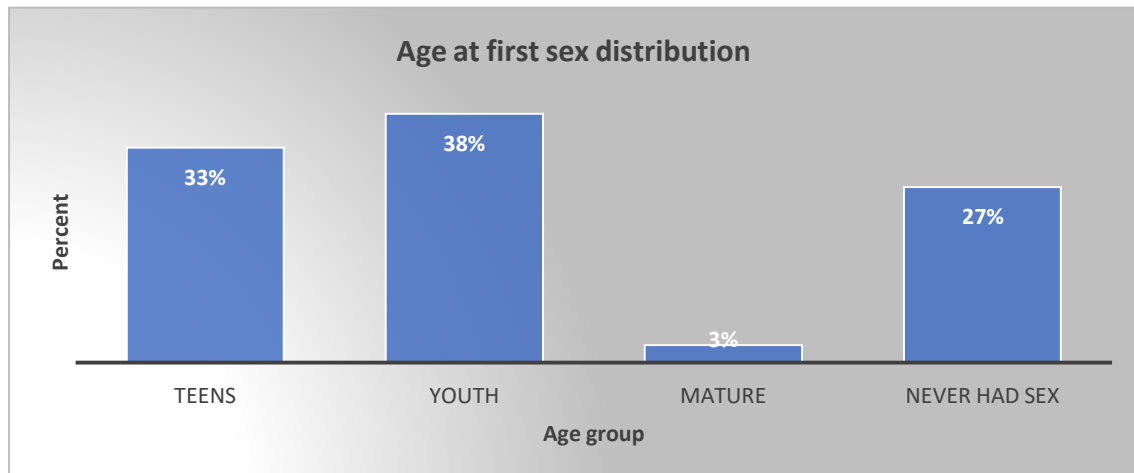
Many respondents are rural residents and accounted for 75% while 25% are urban residents.



4.1.2. Sexual risk behaviours

Age at first sex

Only 3% of the total sample population were mature when they had their first sexual intercourse (above 29 years old), while majority (38%) were youth (20-29) followed by those who were teens (15-19) during their first sexual intercourse. However, 27% never had sexual intercourse.



Male circumcision

Among 13,817 males, more than half (55%) reported not being circumcised while 45% reported being circumcised.

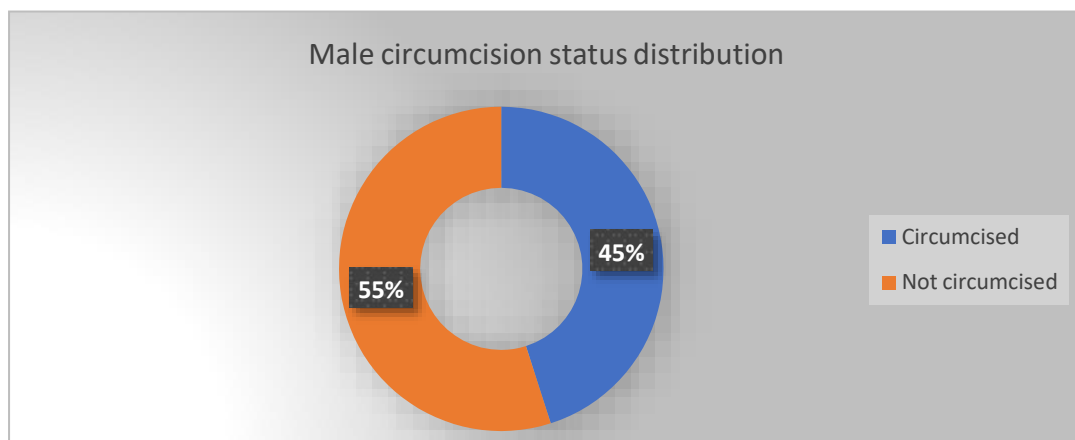


Figure 12. Male circumcision status distribution

Number of sexual partners

Most respondents reported having one sexual partner (55.6%) followed by those who have 2 partners (5.4%). Those who reported having between 3 to 5 sexual partners were 2.5%, and those who have 6 and more accounted for 0.4%. 36.3% are those who do not have any sexual partners.

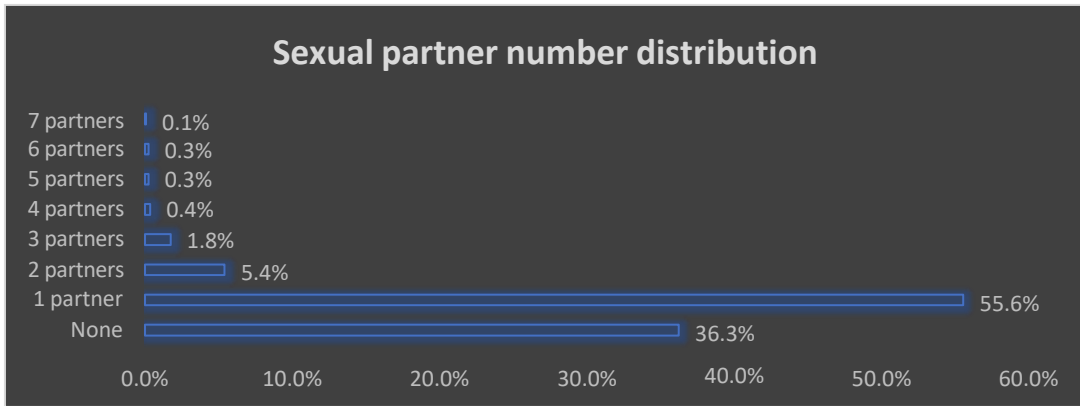


Figure 13. Number of sexual partners distribution

Used condom during the last sexual intercourse

Nine out of ten (90percent) of the sample population did not protect themselves by using condoms during their last sexual intercourse while those that used condoms intercourse accounted for 10%.

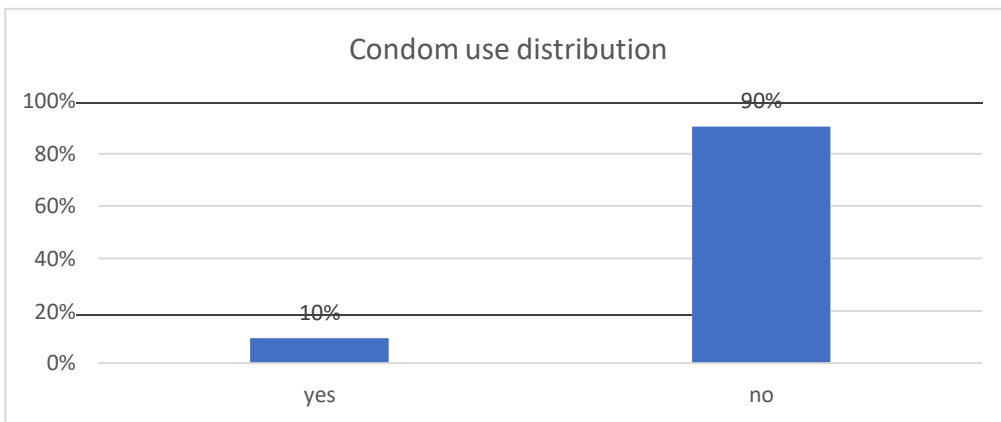


Figure 14. Condom use distribution

Relationship with the last sexual partner

When respondents were asked their relationship with their last sexual partners, 61% reported that they were their partners, 36% reported that they were friends, and only 3% reported that they had sexual intercourse with sex workers.

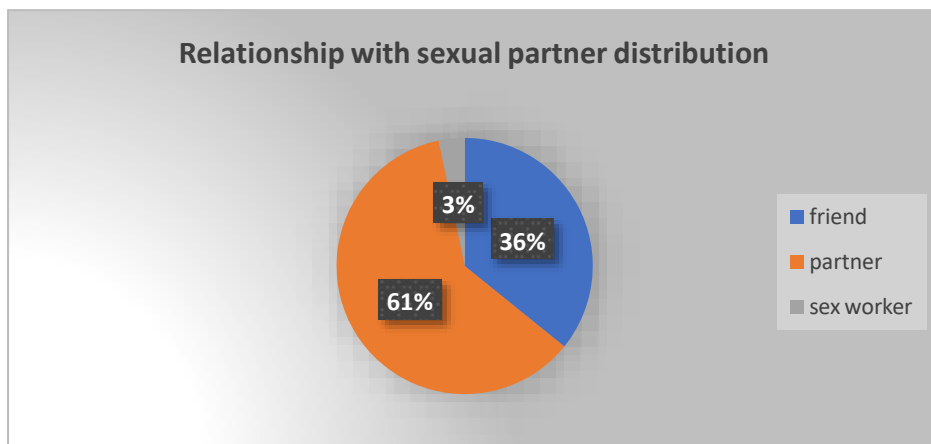


Figure 15. Relationship with sexual partners distribution

Anal sex

Almost all individuals did not engage in anal sex (99.7%). Only 0.3% reported to have engaged in anal sexual intercourse.

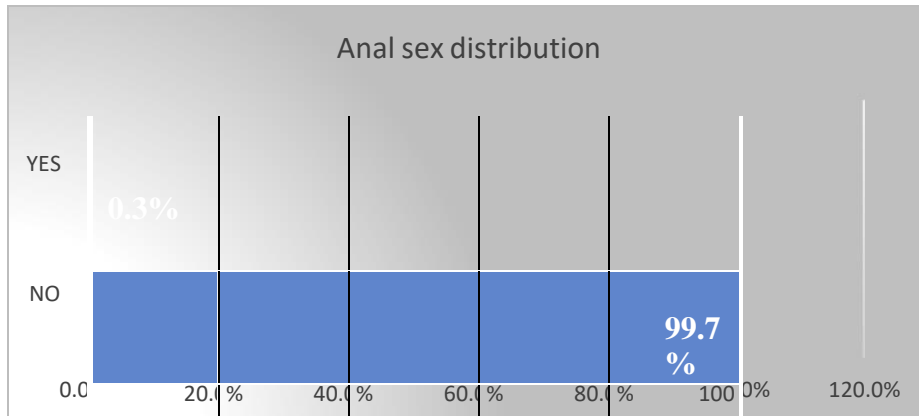


Figure 16. Anal sex distribution

4.2. Distribution of socio-demographic characteristics and sexual risk behaviours by HIV status

Data exploration highlights the association between both socio-demographic characteristics and sexual risk behaviours and HIV status in the sample population. Analysis results revealed that most participants were aged between 15-24 years old (37%). The mean age is 31.6 years old, and the minimum and maximum age were 15 years old and 64 years old respectively. Furthermore, those aged between 25-34 years old, 35-44 years old, 45-54 years old, and 55-64 years old accounted for 26.7%, 18.4%, 10.1%, and 7.8% of the total sample respectively. Respondents in the age group of 35-44 years group had highest rate of HIV positive status (30%) compared to other age groups. The second age group that accounted for higher rate of HIV positive was 25-34 years age group and 45-54 years age group that accounted for 23% followed by 55-64 years old, and 15-24 years old that accounted for 13% and 11% respectively. While for HIV negative status, a shrinking pattern was observed from 15-24 years old (38%) to 55-64 years old (8%).

Most of the sample population was composed with women (55%), and men were 45% of the total sample. Among those who reported being HIV positive, many were women represented at 68% while men were 32%. On the other hand, among those who reported being HIV negative, women were more than men (55% and 45% respectively).

Regarding type of residence, most of the population were rural residents (75%) and 25% were urban residents. Among HIV positive study participants, rural residents were

high represented at 61% and urban residents were 39%, same for HIV negative people, rural population was highly represented at 75% and those who were from urban areas were 25%. This was due to those people who are from rural areas were dominant in the sample population.

Regarding marital status, the majority reported that they are married/living together (48%) followed by those who never married (42%), and 5% of those who are separated. The least reported marital status was divorced (1%) and widowed (4%). HIV prevalence was high among married people (48.9%). Those who never married, widowed, and separated accounted for 19.3%, 16.3%, and 14.8% respectively. The least presented marital status was divorced which might be because they were underrepresented in the sample.

More than half of the sample population had primary level of education (60%) and 26% had secondary level of education. The least reported education level was higher education that accounted for 5%. However, 9% of the sample population had no education. Among HIV positive individuals, 64% had primary level of education, 17% had secondary education, 2% had higher education, and 17% had no education.

Results also showed that the highest percentage of HIV positive person was unemployed (56%) and employed people accounted for 44%. Similarly, many people among HIV negative were unemployed (60%).

Regarding economic status, an increasing pattern was observed from lowest to fifth quintile with 17%, 18%, 19%, 20%, and 26% respectively. For both HIV positive and HIV negative people, respondents from fifth wealth quintile, fourth wealth quintile, and middle wealth quintile were highly represented by 26% HIV negative and 31% HIV positive, 20% HIV negative and 24% HIV positive, 19% and 17% respectively.

Respondents were also asked questions concerning their sexual behaviours which include their age at first sex, consistent condom use, multiple sexual partners, circumcision status, and their attitude towards some statements. Results revealed that most HIV positive people had between 15-19 years old (51.4%) and 20-29 years old (38.5%) when they first had sex, 74.6% previously had sex without using condom, 15.7% had multiple sexual partners, 72.2% men are not circumcised, 90.8% disagreed with the first statement, and 58.3% agreed with the 2nd statement.

Lastly, most of imbalances obtained were due to that some classes were represented by higher percentages than others.

Table 2: Distribution of socio-demographic characteristics and sexual risk behaviours by HIV status

Demographic Characteristics	HIV Negative		HIV Positive	
	Frequency	Percentage	Frequency	Percentage
Gender:				
Female	16,260	55%	632	68%
Male	13,515	45%	302	32%
Age:				
15-24	11,269	38%	103	11%
25-34	7,983	27%	212	23%
35-44	5,368	18%	279	30%
45-54	2,870	10%	218	23%
55-64	2,285	8%	122	13%
Residence:				
Urban	7,331	25%	362	39%
Rural	22,444	75%	572	61%
Marital Status:				
Married/cohabitating	14,338	48%	457	49%
Never married	12,815	43%	180	19%
Separated	1,381	5%	138	15%
Divorced	157	1%	7	1%
Widowed	1,084	4%	152	16%
Education level:				
No education	2,714	9%	157	17%
Primary	17,795	60%	595	64%
Secondary	7,825	26%	160	17%
Higher	1,441	5%	22	2%
Employment status:				
Employed	11,943	40%	411	44%
Unemployed	17,832	60%	523	56%
Wealth quintile:				
Lowest	5,148	17%	143	15%
Second	5,441	18%	124	13%
Middle	5,657	19%	156	17%
Fourth	5,882	20%	225	24%
Fifth	7,647	26%	286	31%
Age at first sex:				
Teens (15-19)	9,579	32%	480	51%
Youth (20-29)	11,274	38%	360	39%
Mature (30-above)	777	3%	30	3%
Never had sex	8,145	27%	64	7%
Condom used last sex:				

Yes	2,712	9%	237	25%
No	27,063	91%	697	75%
Multiple sexual partner:				
1	16,583	56%	493	53%
>1	2347	8%	147	16%
None	10,845	36%	294	31%
Sexual partner relationship:				
Friend	10,780	36%	219	23%
Partner	18,061	61%	649	69%
Sex worker	934	3%	66	7%
Circumcised:				
Yes	6,143	45%	84	28%
No	7,372	55%	218	72%
Attitude 1:				
Agree	3,473	12%	81	9%
Disagree	25,673	86%	848	91%
Don't know	629	2%	5	1%
Attitude 2:				
Agree	13,507	45%	545	58%
Disagree	14,360	48%	341	37%
Don't know	1,908	6%	48	5%

4.3. Predicting new HIV infections using machine learning algorithms

Different algorithms of machine learning were used to find the best algorithm that accurately predicts the model such as logistic regression, random tree forest and gradient boost. To build the train model that predicts HIV infection status, we used the train dataset. The test data was used to test the generalizability/predictability of the model on the data different from the data on which the model was trained. The test set was 28% and the training set was 72% of the dataset. To make decision on the accuracy and the fitness of the model, data must be tested on unseen data. The performance of the model on new data may determine whether the model is fit or not. The unseen data are considered as test set, and the training set consists of existing/available data.

4.3.1. Logistic regression

Logistic regression is supervised learning algorithm used to classify categorical variables (variables with two or more possible outcomes) like yes/no, true/false, or 1/0 (sarvagyaagrawal — May 23, 2021) Logistics regression was used to predict HIV infection status and the accuracy of 0.76 was obtained for both train and test sets. This shows that the model does not overfit and is stable.

- Accuracy train: 0.758732212160414
- Accuracy: 0.7417582417582418

The confusion matrix was built to predict true HIV positives.

At a threshold of 0.5, precision of 74.9%, recall of 64.5%, F1-score of 67.6 were obtained and the model accuracy was at 74.2% which shows that the model was classified at a moderate level with 344 true negatives, 80 false positives, 108 false negatives were, and 196 true positives.

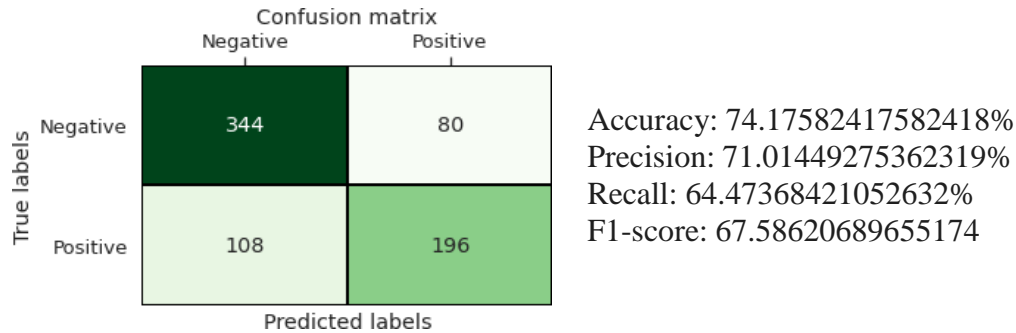


Figure 17. Logistic regression-Confusion matrix at 0.5 threshold

A second confusion matrix was built at a threshold of 0.35 and precision of 62.5%, recall of 79.6%, F1-score of 70 were obtained where the model accuracy was at 71.6%. the predicted values were 279 true negatives, 145 false positives, 62 false negatives, and 242 true positives. This threshold of 0.35 is better than the threshold of 0.5 since it minimizes the false negatives, and it increases the recall and F1-score.

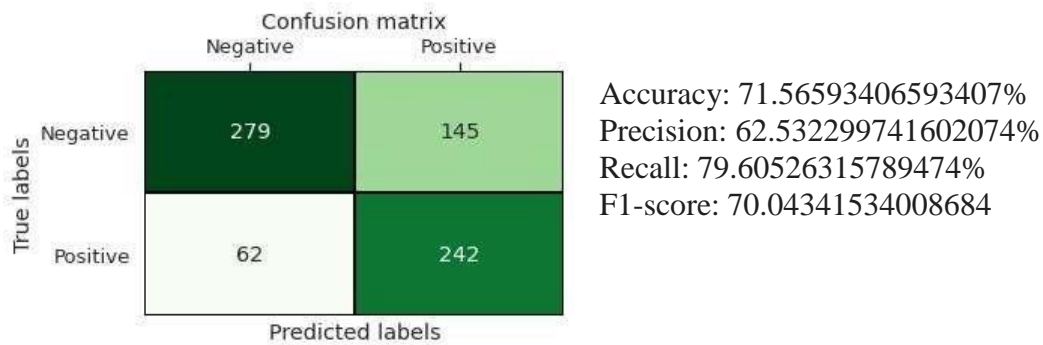


Figure 18. Logistic regression-Confusion matrix at 0.35 threshold

ROC Curve

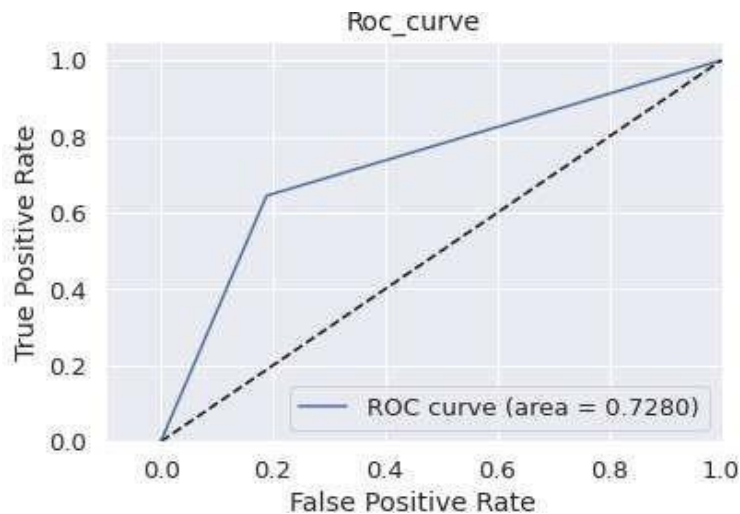


Figure 19. Logistic regression-ROC curve at 0.35 threshold

4.3.2. RANDOM TREE FOREST

With the random tree forest, $n_estimators=10,000$, $max_depth=30$, $max_features=20$, $min_samples_leaf=12$, $random_state=0$; the accuracy of train set, and test set was 0.76 and 0.76 respectively which are similar, this means that the data do not overfit rather fit well.

2 confusion matrixes were built to predict HIV infections.

1. Confusion matrix at 0.5 threshold

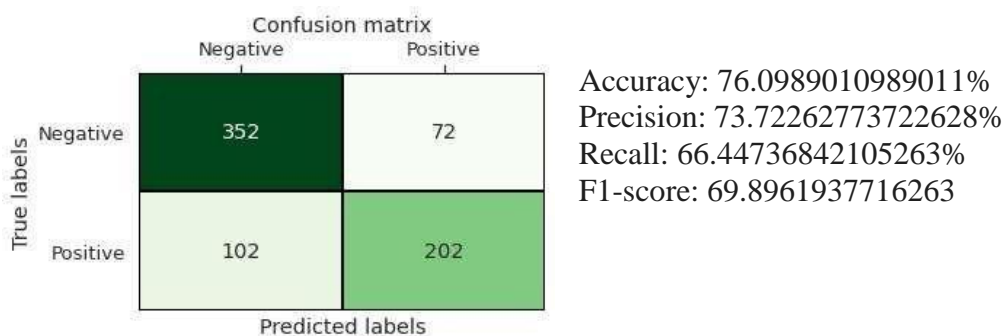


Figure 20. Random tree forest-Confusion matrix at 0.5 threshold.

We acquired an accuracy of 76.1 percent, a precision of 73.7 percent, a recall of 66.4 percent, and an F1-score of 69.9 at the threshold of 0.5. 352 true negatives, 72 false positives, 102 false negatives, and 202 true positives were the predicted figures.

2. Confusion matrix at 0.35 threshold

Our results at the threshold of 0.35 were accuracy of 71.15 percent, precision of 61.2 percent, recall of 84.5 percent, and F1-score of 70.9. According to the predictions, there would be 261 true negatives, 163 false positives, 47 false negatives, and 257 true positives.

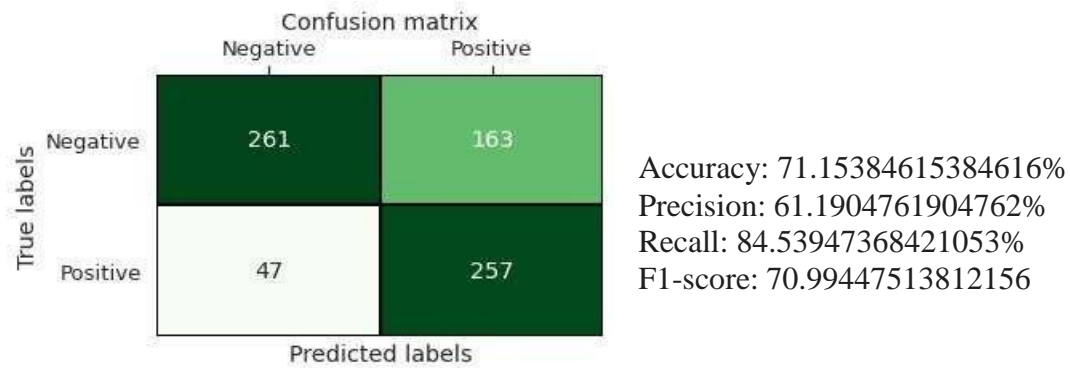


Figure 21. Random tree forest-Confusion matrix at 0.35 threshold.

The confusion matrix predicts well HIV infections at 0.35 threshold since it minimizes the number of false negatives and increases F1-score value and the recall.

ROC C urve

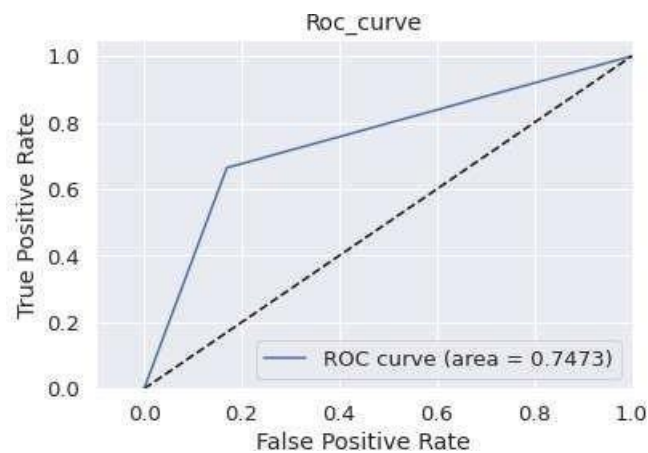


Figure 22. Random tree forest-Roc curve at 0.35 threshold

4.3.3. GRADIENT_BOOST

With the gradient boost, `max_depth=4`, `n_estimators=20`, `random_state=1`, `min_samples_split=2`, `learning_rate=0.09`; the accuracy of train set, and test set was 74.5 and 78.8 respectively which are not similar, this means that the model data overfit.

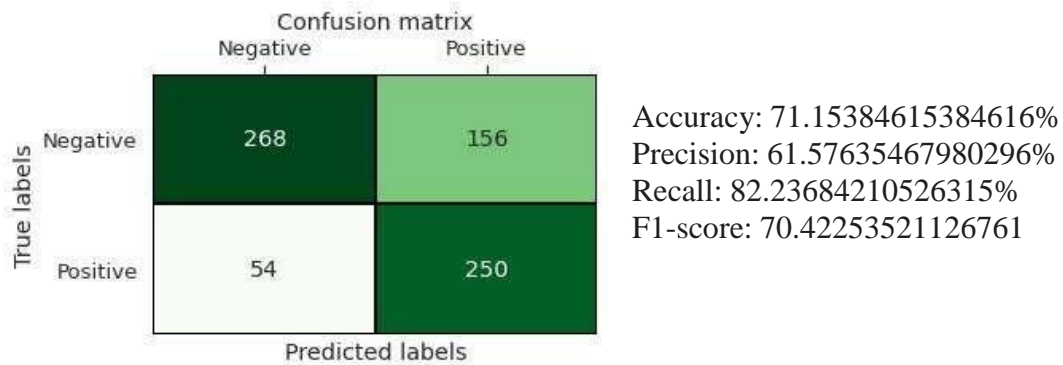


Figure 23. Gradient boost-Confusion matrix at 0.35 threshold.

At 0.35 threshold, the predicted values are 268 true negatives, 156 false negatives, 54 false positives, and 250 true positives with 71.1%, 61.6%, 82.2%, and 70.4 of accuracy, precision, recall, and F1-score respectively.

ROC curve

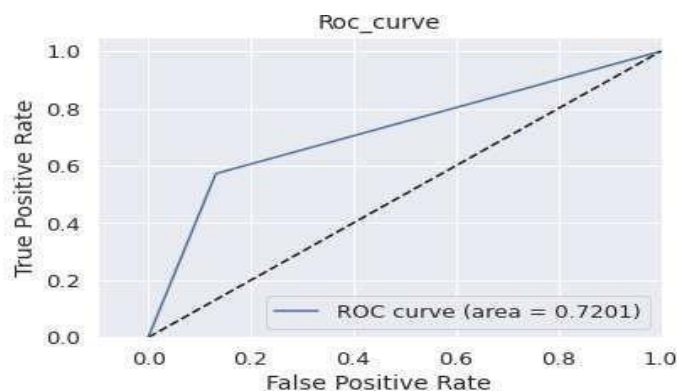


Figure 24. Gradient boost-Roc curve at 0.35 threshold

Since using the gradient boost algorithm, the model data do not fit, this algorithm is not a good classifier for this study given that the area under curve is lower than the one for both logistic regression and random tree forest algorithms.

Chapter 5 DISCUSSION

This chapter discusses the results of the performed analysis in predicting HIV infection among people who practice sexual risk behaviours and highlights the strengths and limitations of the study. It provides recommendations as well to public health practitioners in addressing the problems that lead to HIV infection. In this chapter, an in-depth review and comparison of different machine learning algorithms used is performed and a selection of the one that best predicts HIV infections and allows to disentangle the behavioral factors that contribute to the prediction of HIV infection.

5.1. Comparison of logistic regression, random forest, and Gradient Boost algorithms

For Gradient Boost algorithm, the accuracy of the test set (78.8%) and the accuracy of the training set (74.5%) were not similar which implies that data overfit thus this algorithm is not a good classifier to build a model that predicts HIV infections among people who practice sexual risk behaviours. The area under curve for this algorithm was 0.72.

While for logistic regression, a slight difference between test set accuracy (0.74) and training set (0.76) indicates that data is not perfectly fitted; for random tree forest, the test set accuracy was 0.76 and the training set accuracy was 0.76 as well which showed that using the random tree forest data fit perfectly and the algorithm can be considered as perfect to build model that predicts HIV infections among people with sexual risk behaviours.

Confusion matrixes were built to choose the best algorithm at different thresholds and threshold that reduces mistakes in classification was considered. Using both logistic regression and random forest algorithms, confusion matrixes were compared at 0.5 and 0.35 thresholds and the threshold that reduces the false negatives, increases the true positives, increases the recall and F1-score was 0.35.

For logistic regression at a threshold of 0.5, the accuracy, precision, recall and F1-score were 74.2%, 74.9%, 64.5% and 67.6 were respectively obtained with predicted values of 344 true negatives, 80 false positives, 108 false negatives were, and 196 true positives. While at a threshold of 0.35, the accuracy, precision, recall and F1-score were 71.6%, 62.5%, 79.6% and 70 were obtained, with predicted values of 279 true negatives, 145 false positives, 62 false negatives, and 242 true positives. This threshold of 0.35 is better than the threshold of 0.5 since it minimizes the false negatives, and it increases true positive, the recall, and F1-score. At 0.35 threshold when the roc curve was built, the area under curve value was 0.73

For random tree forest at a threshold of 0.5, we obtained an accuracy of 76.1%, precision of 73.7%, recall of 66.4%, and F1-score of 69.9 and the predicted values were 352 true negatives, 72 false positives, 102 false negatives, and 202 true positives. While at a threshold of 0.35, an accuracy of 71.15%, precision of 61.2%, recall of 84.5%, and F1-score of 70.9 were obtained and predicted values were 261 true negatives, 163 false positives, 47 false negatives, and 257 true positives. 0.35 threshold is better since it minimizes the false negatives, increases true positives, recall and F1-score. At 0.35 threshold when the roc curve was built, the area under curve value was 0.75.

For the test set accuracy and the train set accuracy were similar which indicated that data do not overfit rather perfectly fit, it is better to use the random forest. In addition, random forest algorithm had the greatest value of are under curve (0.75) compared to other algorithms. At 0.35 threshold, using random forest algorithm, highest recall and F-score were found, false positives were greatly reduced, and true positives were greatly increased.

Model and machine learning algorithm selection

As mentioned above, random forest was found to be a better algorithm to predict HIV infection among people that practice sexual risk behaviours.

5.2. Social demographic-sexual risk behaviours-Random tree forest

The factors that affect the model were identified based on feature importance (see figure 25). Among social demographic variables, being in the age group of 15-24, being widowed or single, and having primary level of education were found to be factors that influence the HIV infection. While not having used condoms during last sexual

intercourse, having debuted sex at a younger age (under 20) and having multiple sexual partners (>1) were revealed to be risk behaviours that highly influenced the model that predicts HIV infections.

In this study we tried different algorithms since for example, a study by Oluwabukola (2019) revealed that decision tree does not perform well to predict HIV infection among women of reproductive age in south Africa where the accuracy was 64% and stated that there was overfitting of data accounted for by pruning and resulted in the dataset not capturing the true statistics of the data. (Oladokun, 2019)

In accordance with the study by Oluwabukola (2019) which revealed that women who first had sex at above 18 years old were at lower risk of getting infected with HIV and the study by Munthali and Zulu (2007), Stockl, Karla, Jacobi, and Watts (2013), and Wand and Ramie (2012) found that those who first had sex at below 18 years were most likely to be HIV positive, this study results similarly revealed that sex debut at an early age (18 and below) positively predict HIV infection.

In agreement with the same study by Oluwabukola (Oladokun, 2019) which reported that having tertiary level of education reduces the risk of having HIV, this study revealed that people with lower level of education (primary) greatly influence HIV positivity.

On contrary, this study found that people living in urban areas influenced the HIV serostatus while the study by Oluwabukola (Oladokun, 2019) stated that living in urban areas reduces the risk of getting HIV. This is probably because in urban areas is where people are exposed to sexual reproductive health and rights education. However, living in urban areas can also influence HIV seropositivity since it is where young people are more exposed to sexual risk behaviours like having paid sex for living and most young people who have sex paid have higher chances to not use condoms to earn more money with a limited power to discuss condom use especially that the study revealed that the age difference of between 10-20 years was found to influence HIV serostatus in our model.

In agreement with Johnson et al. (2010), and Kalichman et al. (2007) study that showed that consistent condom use is reduced among people with more than one sexual partner, our study results also found that having multiple sexual partners was observed to

influence HIV serostatus where people with one sexual partner reduces the risk of getting HIV.

The study's findings showed that those who did not use condoms during their most recent sexual encounter had an impact on their HIV seropositivity. Therefore, inconsistent condom use is another important factor that affects HIV serostatus. A study by Kelly Hallman (2005) showed that girls with lower income are subjective to not negotiate condom since they depend on their partners who might exert power on them (Hallman, 2005). In addition, a study by Stephen, Monique, Susan, and Kevin (2014) on attitudinal and behavioral characteristics predict high risk sexual activities in Tanzania revealed that incidence of unprotected sex among youths in Tanzania was high and highest in the age group of 15-19. (Aichele, Mulder, James, & Grimm, 2014).

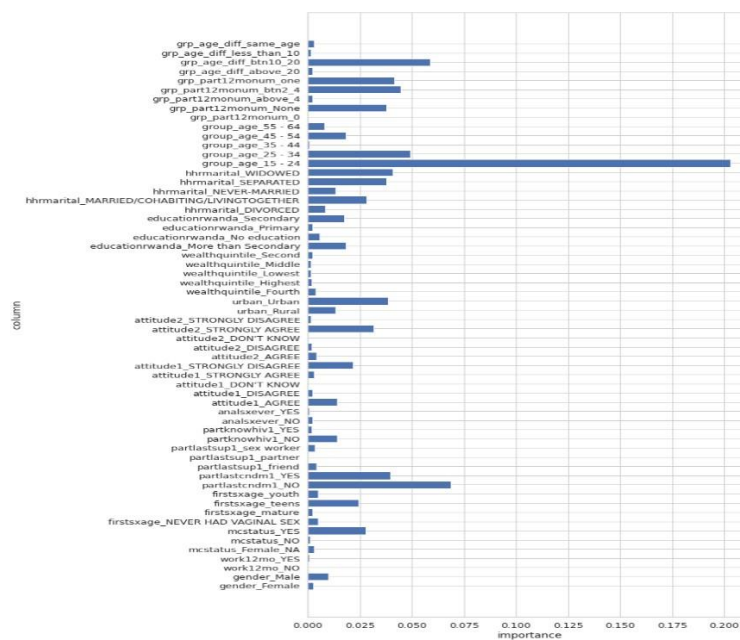


Figure 25. Feature importance techniques

5.3. Limitations of the study

- The study intended to use all sexual risky behaviours; however, we were not able to get data on drug injections.
- Another limitation was that HIV positive study participants were only 3% and this resulted in high skewness to HIV negative.
- Since machine learning techniques do not compute the causal effect, variables may have a strong correlation between themselves but no causality effect and this might lead to misinterpretation of the findings, hence one should be vigilant when interpreting the model.

Chapter 6 CONCLUSION

In summary, the study's overall objective was to build a model that predicts the occurrence of new HIV infections and people with sexual risk behaviours using machine learning techniques and to identify potential risk factors that influence the model that predicts HIV infections among sexual risk behaviours individuals.

HIV is still considered as a global health epidemic and HIV prevalence remains static at 3% in Rwanda with increased prevalence among sex workers and MSM.

Three supervised learning algorithms were used to train HIV data to predict HIV infection among people who practice sexual risk behaviour and those are, logistic regression, random tree forest, and gradient boost. Using the afore mentioned algorithms, random tree forest was found to accurately predict HIV infection with the accuracy of 71.15, F1-score of 70.9 and a recall of 84.5 at a threshold of 0.35 that increases the number of true positives and reduces the number of false positives.

Using the feature importance technique, inconsistent condom use was found to influence the new HIV infection prediction model. In addition, having multiple sexual partners and early age sex debut were found to be sexual risk factors that to influence the new HIV infections prediction model.

Furthermore, the demographic factors found to influence HIV serostatus were being in the age group of 15-24, having lower level of education, being widow and single, and residing in urban areas. Youths in the 15-24 age group and people with low level of education are at more risk of acquiring HIV since they have limited knowledge of HIV prevention and are not confident enough to negotiate safe sex, in addition, they are also exposed to sexual risky behaviours like having paid sex where they might even opt to not use condoms for them to earn more money. This finding is similar to findings by (Aichele, Mulder, James, & Grimm, 2014) whose study revealed that incidence of unprotected sex among youths in Tanzania was highest in the age group of 15-19. In addition, being widow and single since they are mostly not committed to one person and residing in urban areas as well were found to be demographic factors that influence HIV serostatus, and this is mostly because they are more exposed to practice sexual risky behaviours like having paid sex, using drugs, and having multiple sexual partners.

Chapter 7 RECOMMENDATIONS

Inconsistent use of condom was found to be a factor that influence the determination of HIV status especially that as mentioned in the discussion section other studies revealed that people with multiple partners tend to not use condoms as their partners exert power on them and they get scared of negotiating sex, therefore, health public practitioners that are involved in HIV related programs should make emphasis on improving safe sex negotiation skills during HIV prevention and transmission methods education.

As it was observed that early age sex debut also influences HIV status determination, programmes should be focused on educating young, and adolescent children on the prevention of HIV using the nationally approved SRHR curriculum.

REFERENCES

- (WHO), W. H. (2015, January 01). *HIV and Young People who Inject Drugs*. Retrieved from www.who.int: <https://www.who.int/publications-detail-redirect/WHO-HIV-2015.10-eng>
- Aichele, S. R., Mulder, M. B., James, S., & Grimm, K. (2014). *Attitudinal and Behavioral Characteristics Predict High Risk Sexual Activity in Rural Tanzanian Youth*. Canada: Plos One.
- Alem, G., & Teklewoini, M. (2018). Risky sexual behavior practice and associated factors among secondary and preparatory school students of Aksum town, northern Ethiopia, 2018. *Research gate*, 7.
- Avert. (2017). Origin of HIV & AIDS. *Avert*, 5.
- Baral, S., Logie, C. H., Grosso, A., Wirtz, A. L., & Beyrer, C. (2013). Modified social ecological model: a tool to guide the assessment of the risks and risk contexts of HIV epidemics. *BMC Public Health*, 8.
- Cai, Y., Shi, R., Shen, T., Pei, B., Jiang, X., Ye, X., . . . Shang, M. (2010). Research article A study of HIV/AIDS related knowledge, attitude and behaviors among female sex workers in Shanghai China. *BMC Public Health*, 7.
- Crepaz N, M. G. (202). *Towards an understanding of sexual risk behaviour in people living with HIV: a review of social, psychological , and medical findings*.
- Data, U. (2019). *Global information and education on HIV and AIDS*. www.avert.org.
- Hallman, K. (2005). Gendered socioeconomic conditions and HIV risk behaviours among young people in South Africa. *African Journal of AIDS Research*, 37-50.
- Jeffrey D. Fisher, P., & Smith, L. (2010). Secondary Prevention of HIV Infection: The Current State of Prevention for Positives. *National Institutes of Health*, 15.
- Jolly. (2017). *Assessment of risky sexual and practice among Aksum University students*. Shire Town, Tigray, Ethiopia: Awoke Kebede.
- Jolly, D. P. (2012). SEXUAL RISK BEHAVIOUR AMONG HIV-POSITIVE PERSONS IN. *Ghana Medical Journal*, Volume 46, Number 1.
- Kebede, A., B. M., & Gerensea, H. (2017). Assessment of risky sexual behavior and practice among Aksum University students, Shire Campus, Shire Town, Tigray, Ethiopia, 2017. *BMC Research Notes*, 6.
- Leventhal AM, e. a. (2017). *Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence*. JAMA: Jolly, 2017.
- Mathews, C. (2010). *Reducing Sexual Risk Behaviours*.
- McSharry, P. E., Mutai, C., Ngaruye, I., & Musabanganji, E. (2021). Use of Machine Learning Techniques to Identify HIV Predictors for Screening. *BMC Medical Research Methodology*, 12.
- N.M.NCUBE, AKUNNA, J., F.BABATUNDE, NYARKO, A., N.J.YATICH, W.ELLIS, . . . P.E.JOLLY. (2012). Sexual Risk Behaviour Among HIV-Positive Persons in Kumasi, Ghana. *Ghana Medical Journal*, 8.
- Ochonye, B., Folayan, M. O., Fatusi, A. O., Bello, B. M., Ajidagba, B., Emmanuel, G., . . . Jaiyebo, T. (2019). *Sexual practices, sexual behavior and HIV risk profile of key populations in Nigeria*. Lagos, Nigeria: BMC Pubic Health.
- Oladokun, O. O. (2019). *Predicting HIV Status Among Women in South Africa Using Machine Learning: Comparing Decision Tree Model and Logistic Regression*. Johannesburg: Faculty of Humanities, University of the Witwatersrand.

- Orel, E., Esra, R., Estill, J., Marchand-Maillet, S., Merzouki, A., & Keiser, O. (2020). *Machine learning to identify socio-behavioural predictors of HIV positivity in East and Southern Africa*. Geneva.
- Overs, C. (2002). *An analysis of HIV prevention programming to prevent HIV transmission during commercial sex in developing countries*.
- Prof. Chris Beyrer, M., Stefan D Baral, F., Frits van Griensven, P., Steven M Goodreau, P., Prof. Suwat Chariyalertsak, D., Andrea L Wirtz, M., & Prof. Ron Brookmeyer, P. (2012). Global epidemiology of HIV infection in men who have sex with men. *National Institutes of Health Public Access*, 23.
- Rwanda, N. I. (2015). *Rwanda Health Demographic Survey*. Kigali, Rwanda.
- Sex workers, HIV and AIDS. (2019). *Avert*, 18.
- Singh, Y., Narsai, N., & Mars, M. (2013). Applying machine learning to predict patient-specific current CD4 cell count in order to determine the progression of human immunodeficiency virus (HIV) infection. *African Journal of Biotechnology Vol. 12(23)*, 3724-3730.
- Tim, T. N. (2006). *Predicting HIV Status Using Neural Networks and Demographic Factors*. Johannesburg.
- UNAIDS. (2020). *Seizing the moment*. Geneva.
- Yin, L., Zhao, Y., Peratikos, M. B., Song, L., Zhang, X., Xin, R., . . . Qian, H.-Z. (2018). Risk prediction score for HIV infection: Development and internal validation with cross-sectional data from men who have sex with men in China. *Health and Humana Services Public Access*, 18.

APPENDIX

PREDICTION OF HIV INFECTIONS AMONG INDIVIDUALS WITH SEXUAL RISK BEHAVIOURS IN RWANDA USING MACHINE LEARNING ALGORITHMS

ORIGINALITY REPORT

19% SIMILARITY INDEX	15% INTERNET SOURCES	10% PUBLICATIONS	9% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	-----------------------------

PRIMARY SOURCES

1	phia.icap.columbia.edu Internet Source	1%
2	www.medrxiv.org Internet Source	1%
3	www.science.gov Internet Source	<1%
4	Berhanu Teshome Woldeamanue. "Risky sexual behavior and associated factors among high school adolescents in North Shewa zone, Oromia Region, Ethiopia", PAMJ - One Health, 2020 Publication	<1%
5	uac.go.ug Internet Source	<1%
6	Submitted to University of Bedfordshire Student Paper	<1%

7 Hema Sekhar Reddy [Rajula](#), Giuseppe [Verlato](#),
Mirko [Manchia](#), Nadia Antonucci, Vassilios <1 %

[Fanos](#). "Comparison of Conventional
Statistical Methods with Machine Learning in
Medicine: Diagnosis, Drug Development, and
Treatment", [Medicina](#), 2020

Publication

8 www.tshilidzimarwala.com <1 %
Internet Source

9 Catherine Mathews. "9 Reducing sexual risk
[behaviours: theory and research, successes
and challenges](#)", Cambridge University Press
(CUP), 2010

Publication

10 link.springer.com <1 %
Internet Source

11 bmcpublikealth.biomedcentral.com <1 %
Internet Source

12 www.unaids.org <1 %
Internet Source

13	Submitted to London School of Economics and Political Science Student Paper	<1 %
14	Submitted to Mount Kenya University Student Paper	<1 %
15	Submitted to University College London Student Paper	<1 %
	www.rbc.gov.rw	
16	Internet Source	<1 %
17	www.iapac.org Internet Source	<1 %
18	www.rhsupplies.org Internet Source	<1 %
19	phia-data.icap.columbia.edu Internet Source	<1 %
20	Submitted to Middlesex University Student Paper	<1 %
21	ba.one.un.org Internet Source	<1 %

22	nac.org.ls Internet Source	<1 %
23	Submitted to University of Colorado, Denver Student Paper	<1 %
24	medworm.com Internet Source	<1 %
25	journal.kyu.ac.ke Internet Source	<1 %
26	Submitted to Ebonyi State University Student Paper	<1 %
27	aphaih.org Internet Source	<1 %
28	Nidhi Sharma, S. K. Singh, Sudipta Mondal . "Understanding the Factors Affecting the HIV Epidemic in Maharashtra: Application of Proximate Determinants Framework", Sexuality & Culture, 2012 Publication	<1 %
29	Submitted to University of South Florida Student Paper	<1 %
30	www.ncbi.nlm.nih.gov Internet Source	<1 %
31	Submitted to University of Western Australia Student Paper	<1 %

32	www.ajol.info Internet Source	<1 %
33	www.nfi.net Internet Source	<1 %
34	Yonas Tesfaye, Alemayehu <u>Negash</u> , Tsegaye <u>Tewelde Gebrehiwot</u> , Worknesh <u>Tessema</u> , Susan Anand, <u>Gutema Ahmed</u> , Daniel Alemu. "Is There Association between Risky Sexual Behaviors and Depression Symptoms among Youth? A Case of <u>Jimma University</u> Students, Ethiopia", Psychiatry Journal, 2019 Publication	<1 %
35	www.ukessays.com Internet Source	<1 %
36	Submitted to Colorado State University, Global Campus Student Paper	<1 %
37	Geoffrey <u>Setswe</u> . "Systematic reviews of <u>behavioural</u> interventions for reducing the risk of HIV and AIDS: are we getting the evidence?", SAHARA-J: Journal of Social Aspects of HIV/AIDS, 2012 Publication	<1 %
38	businessdocbox.com Internet Source	<1 %
39	fugunnew.blogspot.com Internet Source	<1 %

40	Submitted to Curtin University of Technology Student Paper	<1 %
41	Submitted to Indira Gandhi University, Meerpur, Rewari-123401 (Haryana) Student Paper	<1 %
42	stacks.cdc.gov Internet Source	<1 %
43	Submitted to Anglia Ruskin University Student Paper	<1 %
44	Bridget L. Draper, Zaw Min <u>Oo</u> , Zaw Win Thein, Poe Poe <u>Aung</u> , Vanessa Veronese, Claire Ryan, <u>Myo Thant</u> , Chad Hughes, Mark <u>Stoové</u> . "Willingness to use HIV pre-exposure prophylaxis among gay men, other men who have sex with men and transgender women in Myanmar", Journal of the International AIDS Society, 2017 Publication	<1 %
45	Submitted to La Trobe University Student Paper	<1 %
46	Submitted to University of Witwatersrand Student Paper	<1 %
47	biblio.ugent.be Internet Source	<1 %

48	www.drugsandalcohol.ie Internet Source	<1 %
49	www.esds.ac.uk Internet Source	<1 %
50	Cinzia di Novi, Lucia Leporatti, Marcello Montefiori. "The role of education in psychological response to adverse health shocks", Health Policy, 2021 Publication	<1 %
51	eprints.usm.my Internet Source	<1 %
52	Submitted to People's Open Access Initiative Student Paper	<1 %
53	Submitted to University of Stellenbosch, South Africa Student Paper	<1 %
54	dhsprogram.com Internet Source	<1 %
55	www.tandfonline.com Internet Source	<1 %
56	interaksyon.philstar.com Internet Source	<1 %
57	ift.ee Internet Source	<1 %
58	paa2007.princeton.edu Internet Source	<1 %

59	Submitted to National College of Ireland Student Paper	<1 %
60	Nijatullah Mansoor, Ramesh Chandra Poonia, Debabrata Samanta. "chapter 18 Predictive Analysis of Diabetes Using Machine Learning Algorithms", IGI Global, 2022 Publication	<1 %
61	Udit Pratap, Saurabh Chhabra. "Breast Cancer Prediction using Different Machine Learning Algorithms", 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021 Publication	<1 %
	bmcmedresmethodol.biomedcentral.com	
62	Internet Source	<1 %
63	dspace.unza.zm Internet Source	<1 %
64	reliefweb.int Internet Source	<1 %
65	www.coursehero.com Internet Source	<1 %
66	www.hilcoe.net Internet Source	<1 %
67	www.tvassignmenthelp.com Internet Source	<1 %

68	iiste.org Internet Source	<1 %
69	www.etharc.org Internet Source	<1 %
70	Amjad Alsirhani , Srinivas Sampalli , Peter Bodorik . "DDoS Attack Detection System: Utilizing Classification Algorithms with Apache Spark", 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2018 Publication	<1 %
71	Rafael Vera Cruz de Carvalho , Maria Lucia Seidl-de-Moura , Gabriela Dal Forno Martins , Mauro Luís Vieira . "Culture and developmental trajectories: a discussion on contemporary theoretical models", Early Child Development and Care, 2014 Publication	<1 %
72	Submitted to University of East London Student Paper	<1 %
73	etd.aau.edu.et Internet Source	<1 %
74	hdl.handle.net Internet Source	<1 %
75	journals.plos.org Internet Source	<1 %

76	ukcoalition.org Internet Source	<1 %
77	www.amfar.org Internet Source	<1 %
78	www.hrpub.org Internet Source	<1 %
79	www.research.manchester.ac.uk Internet Source	<1 %
80	www.sundaymail.co.zw Internet Source	<1 %
81	Anne <u>Pithey</u> , Charles Parry. "Descriptive systematic review of sub-Saharan African studies on the association between alcohol use and HIV infection", SAHARA-J: Journal of Social Aspects of HIV/AIDS, 2009 Publication	<1 %
82	Elizabeth <u>Mueni Mutisya</u> , Vincent <u>Muturi-Kioi</u> , Andrew <u>Abaasa</u> , Delvin <u>Nyasani</u> et al. "Feasibility of conducting HIV prevention trials among key populations in Nairobi, Kenya", Research Square Platform LLC, 2022 Publication	<1 %
83	J T F Lau, H Y <u>Tsui</u> , Q S Wang. "Effects of two telephone survey methods on the level of reported risk <u>behaviours</u> ", Sexually Transmitted Infections, 2003 Publication	<1 %

84	<p>Venkatesan Chakrapani, Thilakavathi Subramanian, Pandara Purayil Vijin, Ruban Nelson, Murali Shunmugam, Trace Kershaw. "Reducing sexual risk and promoting acceptance of men who have sex with men living with HIV in India: Outcomes and process evaluation of a pilot randomised multi-level intervention", <i>Global Public Health</i>, 2019</p> <p>Publication</p>	<1 %
85	<p>ir-library.egerton.ac.ke Internet Source</p>	<1 %
86	<p>kb.psu.ac.th Internet Source</p>	<1 %
87	<p>openaccess.city.ac.uk Internet Source</p>	<1 %
88	<p>www.courts.ca.gov Internet Source</p>	<1 %
89	<p>www.lgbtpsychology.org Internet Source</p>	<1 %
90	<p>www.prepwatch.org Internet Source</p>	<1 %

- 91** [Beyrer, Chris, Stefan D Baral, Frits van Griensven, Steven M Goodreau, Suwat Chariyalertsak, Andrea L Wirtz, and Ron Brookmeyer.](#) "Global epidemiology of HIV infection in men who have sex with men", *The Lancet*, 2012.
Publication
-
- 92** [Frank Tanser, Till Bärnighausen, Lauren Hund, Geoffrey P Garnett, Nuala McGrath, Marie-Louise Newell.](#) "Effect of concurrent sexual partnerships on rate of new HIV infections in a high-prevalence, rural South African population: a cohort study", *The Lancet*, 2011
Publication
-
- 93** [Kathleen J. Sikkema.](#) "Mental Health Treatment to Reduce HIV Transmission Risk Behavior: A Positive Prevention Model", *AIDS and Behavior*, 12/15/2009
Publication
-
- 94** [LeVay, Simon.](#) "Human Sexuality", Oxford University Press
Publication
-
- 95** [Li, H., E. Holroyd, X. Li, and J. Lau.](#) "A qualitative analysis of barriers to accessing HIV/AIDS-related services among newly diagnosed HIV-positive men who have sex with men in China", *International Journal of STD & AIDS*, 2014.
Publication
-

96 Stephen R. Aichele, Monique Borgerhoff
Mulder, Susan James, Kevin Grimm. **<1%**
"Attitudinal and Behavioral Characteristics
Predict High Risk Sexual Activity in Rural
Tanzanian Youth", PLoS ONE, 2014
Publication

97 Submitted to Far Eastern University **<1%**
Student Paper

98 Submitted to Higher Education Commission **<1%**
Pakistan
Student Paper

Exclude quotes On

Exclude matches Off

Exclude bibliography On