Brainae Journal of Business, Sciences and Technology
ISSN "2789-374X(print)" "2789-3758(online)"
Volume 22, Issue 2, November 2022
http://www.brainae.org,
info@brainae.org

# THE ANALYTICS AND FORECASTING ALGORITHM ON GENDER-BASED VIOLENCE AND ITS ASSOCIATED FACTORS IN RWANDA.

**BY**
KANYAMUGENGE Clotilde
Reg. No: MSCIT/20/09/6442
**Tel.:** +250 788 575 300
kanyamugengec@gmail.com
**(MASTER OF SCIENCEAND INFORMATION TECHNOLOGY OF THE UNIVERSITY OF KIGALI "UoK")**
**Co-Author:** Dr. MUSONI Wilson
**Tel.:** 0784 186 370
**E-mail:** musoniwilson1@gmail.com
**Received:** 09 November, 2022**; Accepted:** 15 November 2022**; Published:** 21 November 2022
https://brainajournal.com/paper?Id=115

**ABSTRACT:**
*Gender based violence (GBV) is still a common generative health delinquent especially in developing countries of Africa where Rwanda is located. The women are experienced high levels of physical and sexual violence at some point in their lifetime than men. The study objectives determine the prevalence and cases of gender-based violence in Rwanda; establish factors associated with Gender-Based Violence in Rwanda; perform a basic analysis of the GBV Rwandan cases; and to build forecasting model for GBV cases using four models of ML, and interpret the four built models. The population was 500 people from Isange One Stop Center Kacyiru, RWAMUREC and some of the victims of GBV in Rwanda. The stratified and simple randomly sampling techniques were used to select 83 respondents. The questionnaire and documentary techniques were used to collected secondary data. The study used linear regression algorithm, random forest regressor, decision tree regressor and lasso regressor to build predictive data mining models on the Rwandan GBV dataset. Findings on the prevalence of Gender-Based Violence in Rwanda were indicated by tendency of prevalence of GBV in Rwanda which is moderate as confirmed on 62.0%. The associated factors of Gender-Based Violence in Rwanda were revealed by domestic violence and child abuse are among the associated factors of GBV as confirmed by 9.0%; insufficient parental supervision for children victims of GBV, confirmed on rate of 5.0%; peer pressure was stated on rate of 3.0%; drug and alcohol use highly confirmed by 41.0% as main associated factors of GBV in Rwanda; traumatic events is as associated factor of GBV as confirmed by 6.0%; mental illness was confirmed on rate of 2.0%; poverty especially in households are stated as among the associated factors of GBV on rate of 26.0%; while other causes or associated factors unstated in this survey, confirmed on rate of 8.0% in all respondents participated in this survey at Rwanda. Findings on basic analysis of the GBV Rwandan cases showed by the prevalence of GBV crimes during four consecutive years. Findings indicated high rate of child defilement in Eastern province followed by southern province and third was western province in child defilement. Spouse harassment was high in southern province than other provinces in Rwanda. The prevalence of GBV crimes has been depicted at the district level, where the following plot reveal it clearly. Findings indicated GBV per district where Bugesera District presented high rate of child defilement than other districts of Rwanda. The predicted results of the row data which have been randomly picked from the entire dataset where the predicted results of the row data which have been randomly picked from the entire dataset.*
**Key words:** *Gender-Based Violence, associated factors; analytics, forecasting*

## INTRODUCTION

The Gender-based violence is a sad reality in Rwanda. Reports have revealed rape of children and adults, beaten and injured women, and, and murder. Out of the 30 districts of the country, Rwanda ranked first followed by Nyarugenge and Kicukiro as the districts most experiencing violence. This adds to the fact that more than one-third of women (31 percent) in Rwanda have suffered from physical violence since the age of 15 years. In 19 percent of these cases, women had suffered from acts of violence within the last 12 months.

GBV examples include early forced marriages, sexual abuse, infanticide, physical abuse, emotional or psychological abuse, harassment or intimidation, neglecting and abandoning a child, just to name a few. The main GBV forms have been identified in Rwanda included by sexual violence, physical violence, economical violence and psychological violence (Richters, A., Rutayisire,et al., 2018).

Rwanda, like with other nations throughout the world, is fighting to end DGBV and all other types of GBV.  A victim has two options for reporting their incident: they can travel to the Isange One-Stop Center, where they receive integrated services such as medical attention, psycho-social help, and legal support, or they can phone the hotline number (3512), where some officers have received training on GBV.

Although above measures are taken and implemented, GBV is still a major challenge that our society is facing and even the globe. Gender-based violence (GBV) is an important concern that governments and societies must confront with all available resources. This necessitates sufficient planning in order to optimize both resources and budget, which necessitates a detailed grasp of the issue's extent as well as a study of its past impact in order to predict future onset.

Child defilement received cases have been 5,292, where 5,116 is a number of male perpetrating violence against female children. The same as Harassment of a spouse, where 3,016 cases have been received and all 2,837 are men who exercised violence against their female partners. Rape cases have been 1,143, and 1,106 are male who committed it against female (Quarterly & Report, 2021).

All these cases said above, are the GBV cases and have been prosecuted, but the problem remains persistent, as this report mentions. The problem persists and even increases by considering the NPPA report. In 2016-2017, total GBV case have been 3,130, in 2017-2018 cases have increased to 4,592, in 2018-2019, cases have increased to 5,563, as well as in 2019-2020 where cases reached the number of 7,004. Finally, the last report of 2019-2020 shows that GBV cases have been 7,004. This report is depicting how number of cases increases yearly and have been increasing three times during five years, which is a significant nuisance of Rwandan society. (Quarterly & Report, 2021).

## PROBLEM STATEMENT

The factors associated with Gender-Based Violence are significant to women, these are the age, job employment opportunity, having a partner, religion, ethnicity, living arrangement, having a roommate with a boyfriend, and monthly pocket money, all of that are associated factors of GBV (Alemu Basazin & Tadesse, 2021).

According to Rwanda Investigation Bureau statistics (2020) indicated that number of GBV cases has increased to 19.6% in the year 2020 compared to the cases received in the previous fiscal year of 2019 where in that year, RIB received 10 842 GBV related crimes; the highest rate of these cases being child defilement, assault and domestic abuse. Findings also revealed that there are still gaps in the family and community levels in terms of prevention of GBV and its associated factors in Rwanda, because children are at higher risk of sexual abuse than adults especially in high schools in Rwanda. (RWAMREC survey, 2017).

Rwanda ranked on the first followed by Nyarugenge and Kicukiro as the Districts most experiencing violence (MIGEPROF, 2021).  It is therefore, the researcher is motivated to undertake the analytics and forecasting Algorithm on gender-based violence and its associated factors in Rwanda: using data relating to GBV from Rwanda.

## OBJECTIVE OF THE STUDY

The following specific objectives were fulfilled during this study:

1. To determine the prevalence of gender-based violence in Rwanda.
2. To perform a basic analysis of the GBV Rwandan cases.
3. To build forecasting model for GBV cases using four models of ML, and interpret the four built models.
4. To establish factors associated with Gender-Based Violence in Rwanda.

## RESEARCH QUESTIONS

1. What is the prevalence of Domestic gender-based violence in Rwanda, Rwanda?
2. Which are the factors associated with Domestic Gender-Based Violence in Rwanda, Rwanda?
3. How is the DGBV Rwandan dataset described?
4. What should be the future picture of Rwandan DGBV cases; and how the built forecasting models of DGBV cases are assessed?

## CONCEPTUAL REVIEW

This sub-section illustrates an overview of GBV and its associated factors.

### Gender-Based Violence (GBV)

Gender-based violence (GBV) stays as violence this is directed at an individual centered on his or her biological intercourse or gender bibliography. It accommodates physical, emotional, sexual, verbal, and psychological abuse, threats, pressure, and financial or academic deprivation, whether or not taking place in public or non-public life (Women to Women International, 2021).

### Causes of Gender-Based Violence

Gender-based violence in communities cannot be defined by a single factor; instead, a wide range of variables interacts to cause the issue, which means that many determinants lead to GBV. These involve cultural issues, legal implications, financial implications, political concerns, school performance, alcohol consumption, and familial relations. (UNIFEM, 2008).

### Consequences of Gender-Based Violence

(Gender based violence remains affecting humans' rights, annoying their social life and development. It affects them in their households and within the community, it has some consequences such physiological distress and has effects on survivors 'emotions, behaviors as well as their physicality (Jewkes, R., et al., 2019).

GBV also can affect girls' reproductive health and those who have been sexually abused are more susceptible to unplanned pregnancy, HIV infection, and other STDs, mental health-over a third of raped women experience some disorders such as PTSD and when it remains for a long time without being treated it can cause depression, substance abuse even suicidality (Jewkes, R., *et al.,* 2019).

## THEORETICAL FRAMEWORK

### Social Learning Theory

In step with Miller, Patricia H., (2016) social getting to know idea is a principle of getting to know process and social behavior which proposes that new behaviors can be received by means of looking at and imitating others.

The cognitive process takes vicinity in a social context and may occur basically via remark or direct instruction, even in the absence of motor reproduction or direct reinforcement. The learning happens through the remark of rewards and punishments, a process referred to as vicarious reinforcement.

When a selected conduct is rewarded frequently, it maximum in all likelihood persists; conversely, if a selected behavior is constantly punished, it most in all likelihood desists. The principle expands on traditional behavioral theories, wherein conduct is ruled totally via reinforcements, by putting emphasis at the critical roles of diverse internal methods within the learning character.

Social gaining knowledge of concept requires that people can study new behaviors by way of gazing others. The social gaining knowledge of emphasizes the reciprocal courting between social characteristics of the environment, how they may be perceived via individuals, and how prompted and capable someone is to reproduce behaviors they see happening round them.

## CONCEPTUAL FRAMEWORK

Social cognitive learning theory (SCLT) builds a structure for learning, forecasting, and modifying human behavior. The SCLT focuses a strong emphasis on cognitive theories, which analyze how adults and children logically perceive social events, and how these cognitions influence behavior and development**.** (Miller, Patricia H., 2016). The researcher established the relationship between independent variable in terms of GBV and the dependent variable in terms of associated factors of GBV as figure 1 below.

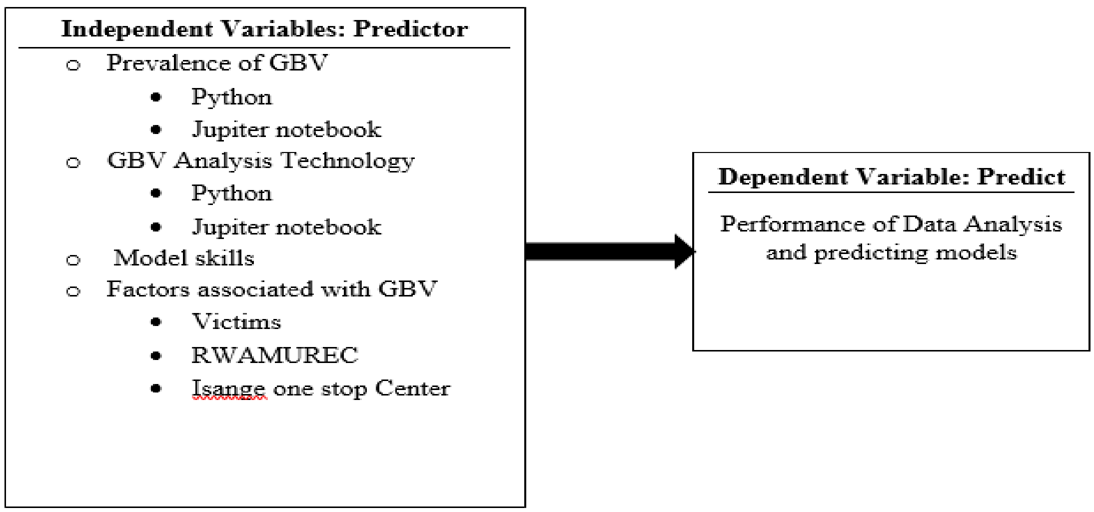


**Figure 1: Conceptual framework**
**Source:** Researcher Conceptualization (2022)

## RESEARCH METHODOLOGY

The study applied qualitative and quantitative approaches. Machine learning was used as a technique for teaching machines to better handle data. A human being may be unable to interpret the pattern or extract information from the data after examining it. In this scenario, a human being uses machine learning.

Target population was 59 employees in charge of GBV at Isange One Stop Center Kacyiru, in Rwanda, 31 people from RWAMUREC in Gasabo, and targeted 410 victims of GBV in Rwanda. That means total target population is 500 people from Isange One Stop Center Kacyiru, RWAMUREC and some of the victims of GBV in Rwanda. The study applied the formulation of Taro Yamane (1982) to control sample size of this study.

Where: $n = \frac{N}{1 + N(e)^2}$

n = Sample Size N = Study Population
e = Margin of error

$$\boldsymbol{n} = \frac{500}{1 + [500*(0.1)^2]} = 83$$

The stratified and simple randomly sampling techniques were used to select 83 respondents as sample size from Isange One Stop Center Kacyiru, RWAMUREC and some of the victims of GBV in Rwanda.
Data-collection instruments" means tests, questionnaires, inventories, interview schedules or guides, rating scales, and survey plans or any

other forms which are used to collect information on substantially identical items from 10 or more respondents.
The secondary data about GBV have been mainly used for analytics and model building which have been provided by NPPA and RIB. The questionnaire was distributed to 83 respondents, and it was composed by close end and open questions. The researcher was expecting the participation rate of 100% for responding to the questions. The five likert scales were used to elicit opinions and perceptions of respondents.

Documentary technique were used by the researcher to obtain the secondary information about a phenomenon where wishes to study. The documents targeted were available reports showing prevalence and cases of GBV from Isange One Stop Center Kacyiru, RWAMUREC in Rwanda. This study used a method that starts with the formulation of research questions, then moves on to the identification of the dataset and a complete examination of the dataset. Based on analytics of different metrics, this study identified the most performing predictive algorithms of machine learning.

The study used linear regression algorithm, random forest regressor, decision tree regressor and lasso regressor to build predictive data mining models on the Rwandan GBV dataset. A statistical indicator is the representation of statistical data for a specified time, place or any other relevant characteristic, corrected for at least one dimension so as to allow for meaningful comparisons.
The statistical indicators were adopted are descriptive indicators (i.e.: frequencies, percentages, etc.); inferential statistical indicators (i.e.: Predictor for GBV and predict for performance of data analysis and with predicting models); performance indicators (i.e.: a quantifiable measure of performance over time for a specific objective for GBV prevention) and with quality indicators. Machine learning is a technique for teaching machines to better handle data.

**DATA INTERPRETATION AND RESULTS**

All the processes required for the feature engineering for the data analysis and forecasting algorithm in the Data Science, which covered the Data cleaning, categorical encoding and checking the null values, which would lead the models' algorithms to not understand the data in categorical type. Furthermore, it embraces the data visualization where multivariate graphs have been found together with data training and data testing to help for determination of the algorithm which best fits with the dataset. A visual representation of the development process, from data preparation to model construction, is shown in the figure.

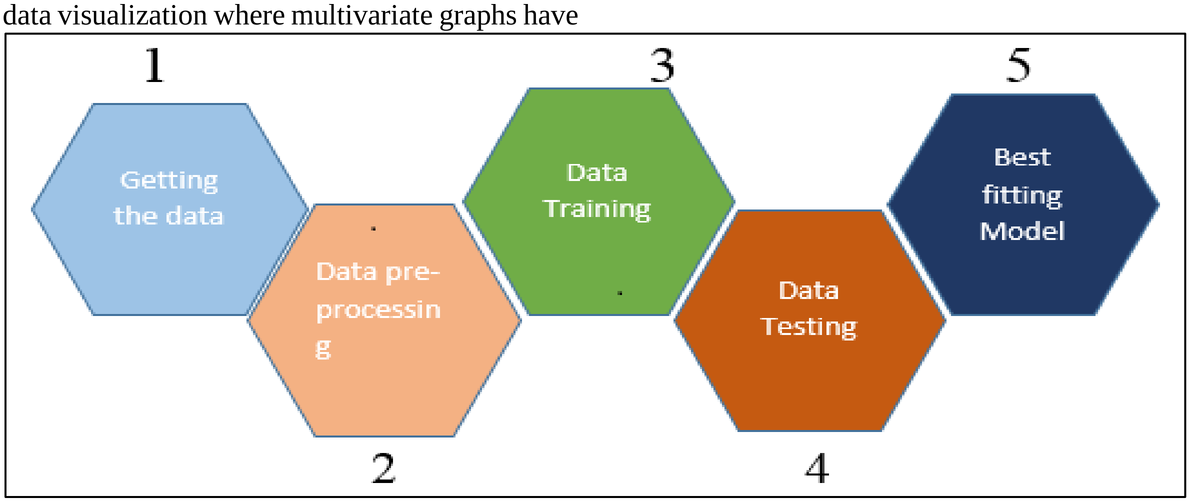


*Figure 2: The workflow of a machine learning process*

**Data Cleaning**

The first step in any data science (DS) or machine learning (ML) workflow is data cleaning. Without clean data, it is much harder to understand the parts of your investigation that really matter. Once the ML model starts training, it becomes unnecessarily difficult to train. Most importantly, to get the most out of your dataset, you need to clean it. In the context of data science and machine learning, data cleansing means filtering and modifying data to make it easier to explore, understand, and model. Filter out parts you do not want or do not need so you do not have to view or edit them. Fix the necessary parts so that they can be used properly. The dataset used needed the data pre-processing to be considered clean.

**Categorical Encoding**

Despite the fact that the majority of machine learning algorithms only work with numerical data, this thesis used categorical data as dataset, where features are not numeric but categorical. These categorical features are taken on levels or values where they can be categorized in a number of ways, such as by age, such as by gender, District or province. A number of machine learning algorithms do not handle categorical data, therefore converting those in numerical values become a necessity. In this thesis, binary encoding has been used to handle the dataset.

**Binary encoding**

Binary encoding is a technique used to convert category data into numerical data by encoding categories as integers and then translating them into binary digits. The figure below is the extract of binary encoded categorical data of the current GBV dataset.

***Table 1:** **Image Table of binary encoded data***

| CRIME_CATEGORY | GENDER | YEAR | AGE_CATEGORY | PROVINCE | number_crimes | CHILD DEFILEMENT | RAPE | SPOUSE HARASSMENT | F | M | Adult | Child | Eastern Province | Kigali City |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RAPE | M | 2021 | Adult | Eastern Province | 3 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| RAPE | M | 2020 | Adult | Kigali City | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| CHILD DEFILEMENT | F | 2023 | Child | Southern Province | 715 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| CHILD DEFILEMENT | M | 2022 | Child | Kigali City | 11 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| SPOUSE HARASSMENT | M | 2020 | Adult | Kigali City | 2 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |

***Source:** Python binary encoded data*

**Data visualization**

In exploratory data analysis, finding interesting patterns depends heavily on data visualization. However, the sheer volume of potential data projections displaying various characteristic subsets that the data analyst must assess makes its application challenging. Data visualization is an essential tool in data analysis since it enables us to visually detect complex structures and patterns in the data.

**1. Multivariate plots**

The first developed plot shows the prevalence of GBV crimes during four consecutive years; from 2019 to 2022.

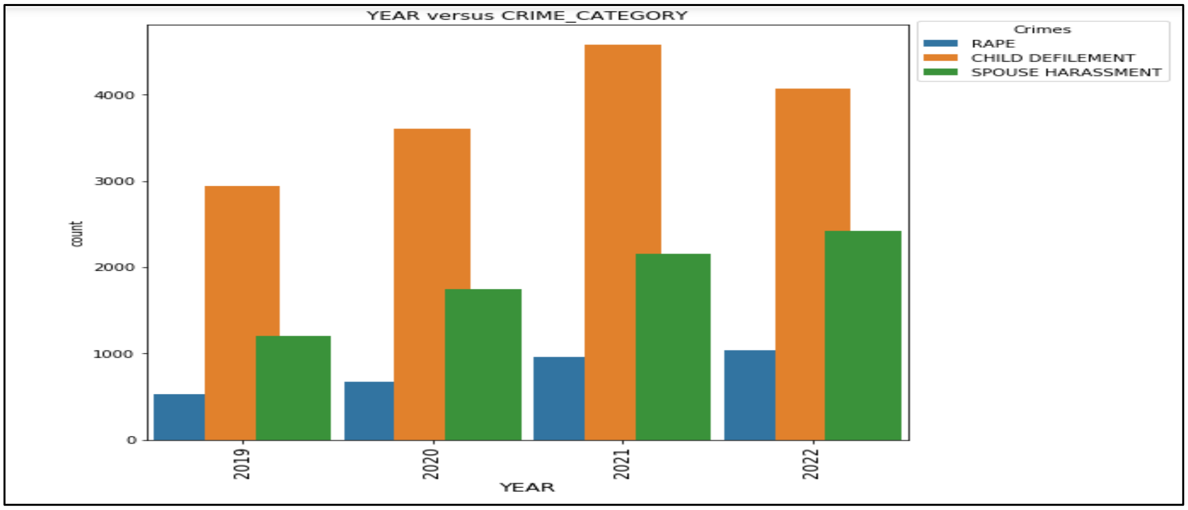


***Figure 3: Plot illustrating the status of GBV crimes in four years***

Findings in figure 2 indicated high rate of child defilement in all years, more especially in the period of 2020 to 2021, where the country as well as the whole globe was facing the covid-19 pandemic with multiple lockdowns, which have been the prior cause to the increase of DGBV crimes in general. The second plot which has been developed shows how GBV crimes are illustrated in four provinces of Rwanda.

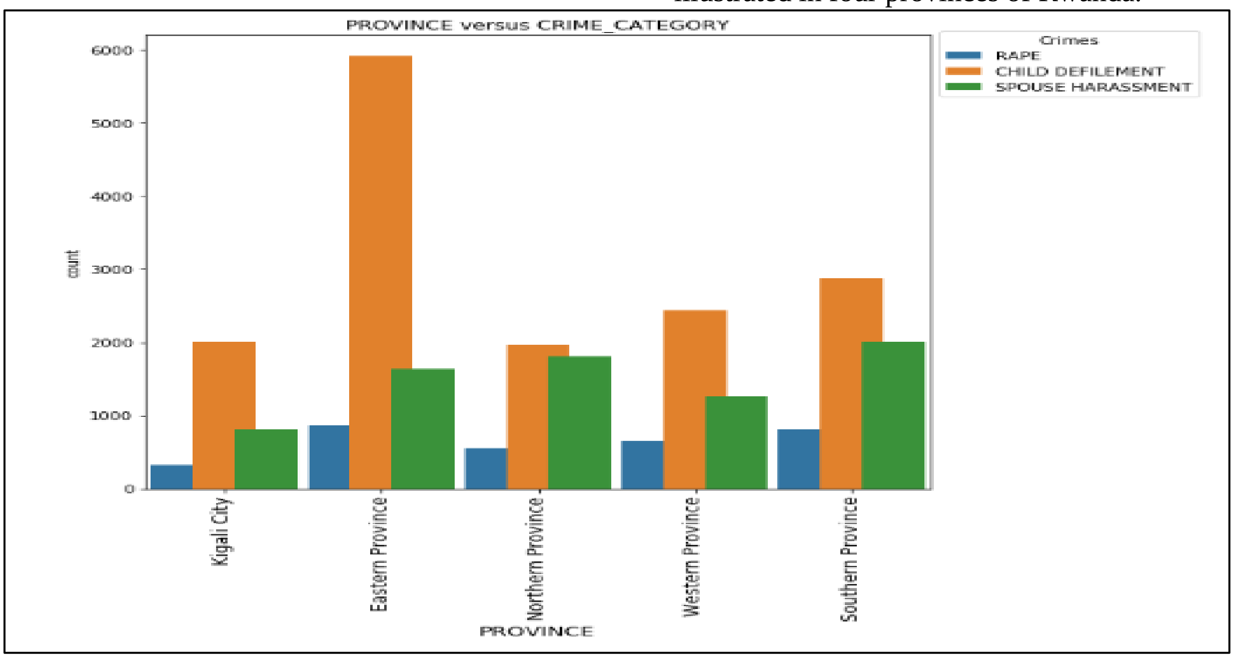


**Figure 4: Plot illustrating the outlook of crimes in all Rwandan provinces**

The high rate of child defilement in Eastern province followed by southern province and third was western province in child defilement. Spouse harassment was high in southern province than other provinces in Rwanda. Apart from provinces, also the prevalence of GBV crimes has been depicted at the district level, where the following plot reveal it clearly.

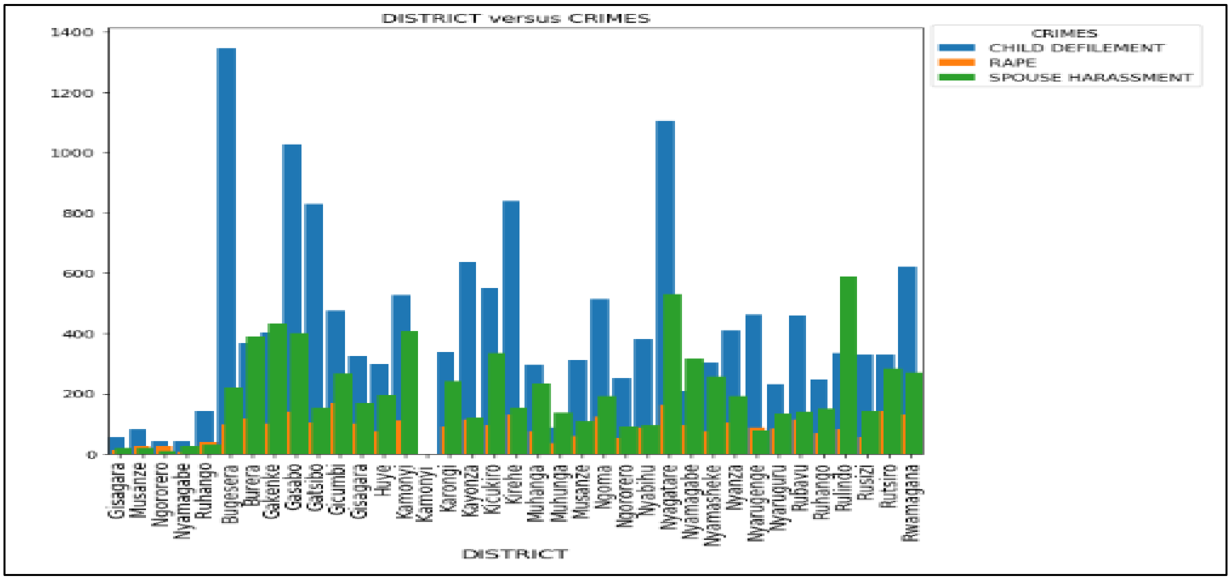


**Figure 5: Plot illustrating the outlook of crimes in all Rwandan districts**

GBV per district where Bugesera District presented high rate of child defilement than other districts of Rwanda and Rulindo district as the first for spouse harassment. The plot below presents how age categories are affected by the GBV crimes, where the following figure clearly demonstrates it.

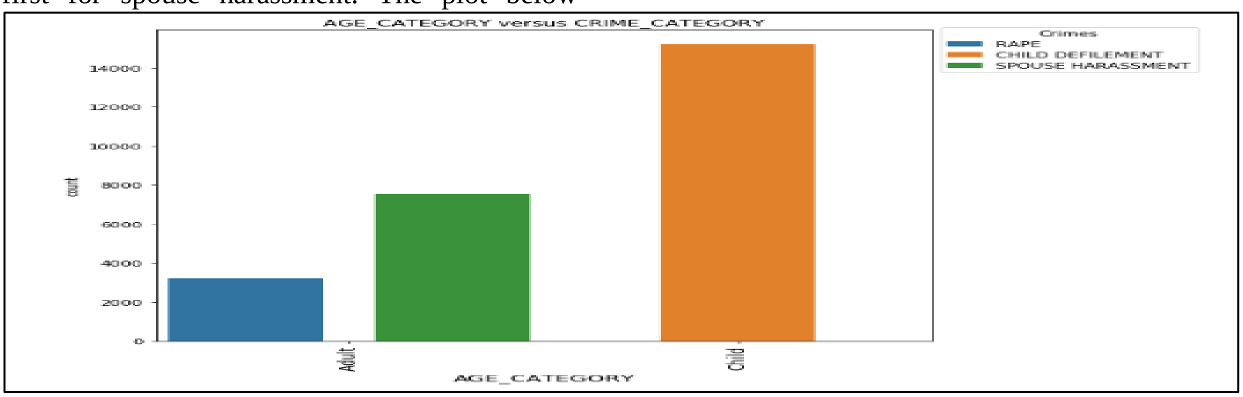


**Figure 6: Plot illustrating the outlook of GBV crimes in different age categories**

GBV by age category, where the results analysis confirmed that child defilement is still high in childhood stage while spouse harassment and rape is mainly for adults.

**2. ML Models results**

**Model 1: LINEAR REGRESSION**

The following results show the values obtained from three metrics used to evaluate the linear regression model.

**R^2 score**:

| Training | Testing |
|---|---|
| 0.62 | 0.62 |

**Mean Square Error (MSE)**:

| Training | Testing |
|---|---|
| 39865045 | 30569.13. |

**Root Mean Square Error (RMSE)**:

| Training | Testing |
|---|---|
| 199066 | 174.84 |

**Mean Absolute Error:**

| Training | Testing |
|---|---|
| 144.82 | 128.84 |

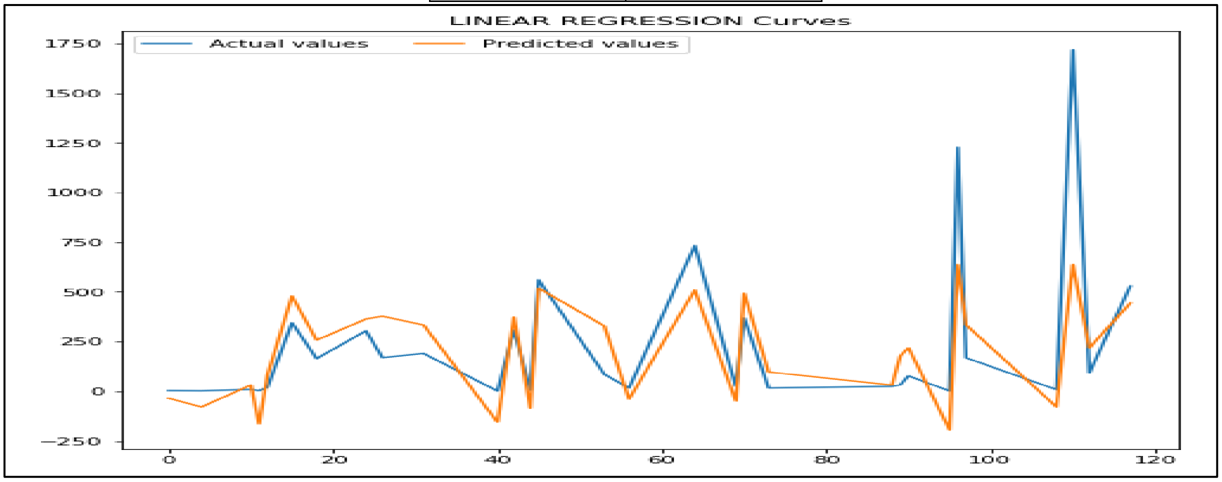


**Figure 7: Linear Regression Prediction Model**

The above curve shows that the Linear Regression model did not fit the data, therefore leads to poor performance. The following are the predicted results of the row data which have been randomly picked from the entire dataset.

**Table 2: Image Table of Linear Regression Model predicted values**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Random Row Data | 0.000000 | 4.000000 | 10.000000 | 11.000000 | 12.00000 | 15.000000 | 18.000000 | 24.000000 | 26.000000 | 31.000000 |
| Actual Values | 3.000000 | 2.000000 | 9.000000 | 3.000000 | 19.00000 | 345.000000 | 164.000000 | 304.000000 | 169.000000 | 190.000000 |
| Predicted Values | -35.780896 | -78.983583 | 30.999207 | -164.978994 | 79.74741 | 479.711699 | 256.968182 | 363.321691 | 376.991482 | 333.788795 |

***Source:*** *Python LRM prediction results*

**Model 2: LASSO**

The following results show the values obtained from three metrics used to evaluate the Lasso model.

**R^2 score**:

| Training | Testing |
|---|---|
| 0.61 | 0.62 |

**Mean Square Error** (**MSE**):

| Training | Testing |
|---|---|
| 41420.53 | 30495.20 |

**Root Mean Square Error (RMSE**):

| Training | Testing |
|---|---|
| 203.52 | 174.62 |

**Mean Absolute Error:**

| Training | Testing |
|---|---|
| 140.53 | 118.84 |

The graph shown below represents the performance of Lasso Algorithm;

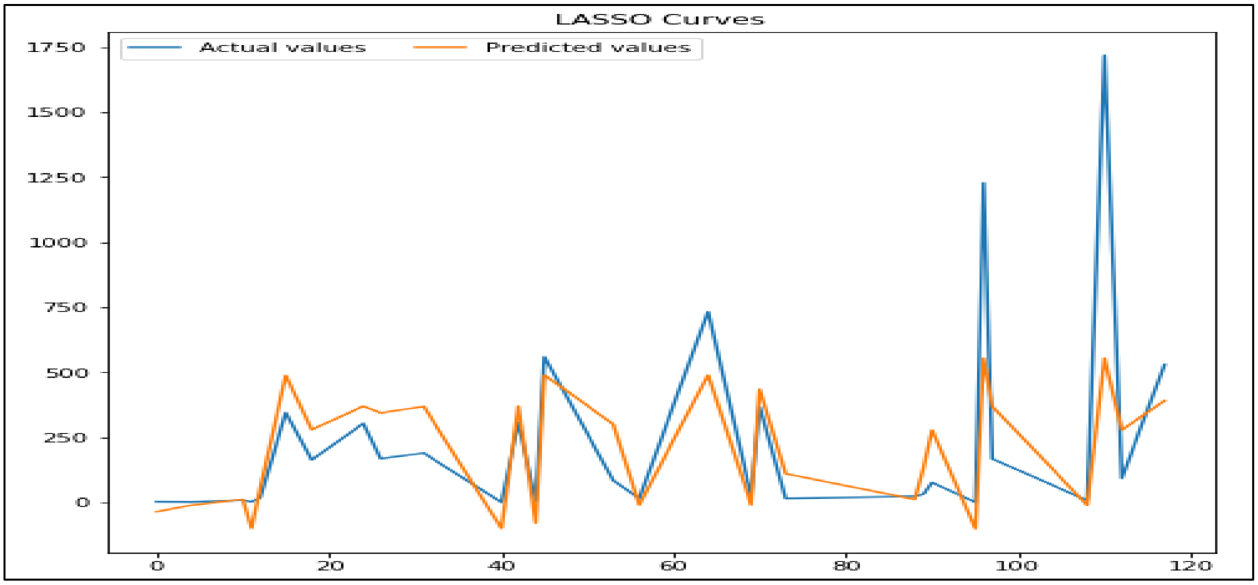


**Figure 8: Lasso Prediction Model**

The above curves show that the also Lasso model did not fit the current GBV dataset, therefore leads to poor performance. The following are the predicted results of the row data which have been randomly picked from the entire dataset.

**Table 3: Image Table of LASSO Model predicted values**

| Random Row Data | 0.000000 | 4.000000 | 10.000000 | 11.000000 | 12.000000 | 15.000000 | 18.000000 | 24.000000 | 26.000000 | 31.000000 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual Values | 3.000000 | 2.000000 | 9.000000 | 3.000000 | 19.000000 | 345.000000 | 164.000000 | 304.000000 | 169.000000 | 190.000000 |
| Predicted Values | -34.875787 | -10.053696 | 12.161087 | -99.309068 | 55.692199 | 488.384718 | 279.837419 | 370.405405 | 344.270699 | 369.092791 |

**Source:** Python LASSO prediction results

**Model 3: Gradient Boosting Regressor**

The following results show the values obtained from three metrics used to evaluate the Gradient Boosting Regressor model.

**R^2 score**:

| Training | Testing |
|---|---|
| 0.74 | 0.77 |

**Mean Square Error** (**MSE**):

| Training | Testing |
|---|---|
| 27262.82 | 18642.99 |

**Root Mean Square Error** (**RMSE**):

| Training | Testing |
|---|---|
| 165.11 | 136.53 |

**Mean Absolute Error:**

| Training | Testing |
|---|---|
| 118.02 | 98.38 |

Similarly, the obtained curves show that the Gradient Boosting Regressor model did not fit the data, therefore leads to poor performance.

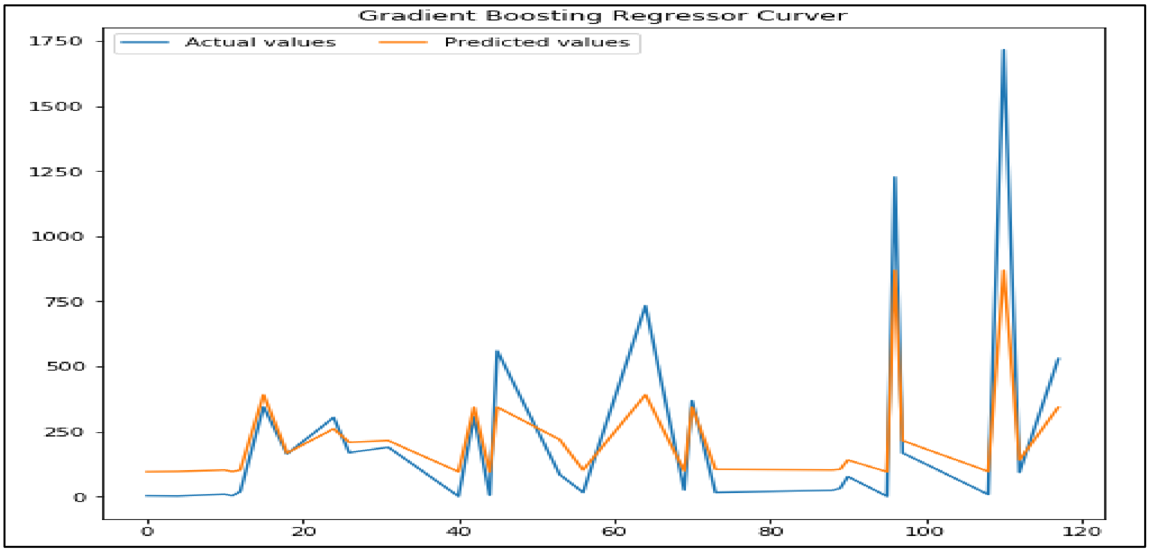


**Figure 9: Gradient Boosting Regressor Model**

The predicted results of the row data which have been randomly picked from the entire dataset.

**Table 4: Image Table of GBR Model predicted values**

| Random Row Data | 0.000000 | 4.000000 | 10.000000 | 11.000000 | 12.000000 | 15.000000 | 18.000000 | 24.000000 | 26.000000 | 31.000000 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual Values | 3.000000 | 2.000000 | 9.000000 | 3.000000 | 19.000000 | 345.000000 | 164.000000 | 304.000000 | 169.000000 | 190.000000 |
| Predicted Values | 95.179999 | 96.395944 | 102.058831 | 95.843468 | 102.058831 | 391.326469 | 167.675396 | 260.849868 | 208.581805 | 215.178453 |

***Source:*** *Python GBR prediction results*

**Model 4: RANDOM FOREST REGRESSOR**

The following results display the values obtained from three metrics used to evaluate the Random Forest Regressor model.

**R^2 score**:

| Training | Testing |
|---|---|
| 0.92 | **0.94** |

**Mean Square Error** (**MSE**):

| Training | Testing |
|---|---|
| 7513.05 | 4413.85 |

**Root Mean Square Error** (**RMSE**):

| Training | Testing |
|---|---|
| 86.67 | 66.43 |

**Mean Absolute Error:**

| Training | Testing |
|---|---|
| 44.24 | 44.44 |

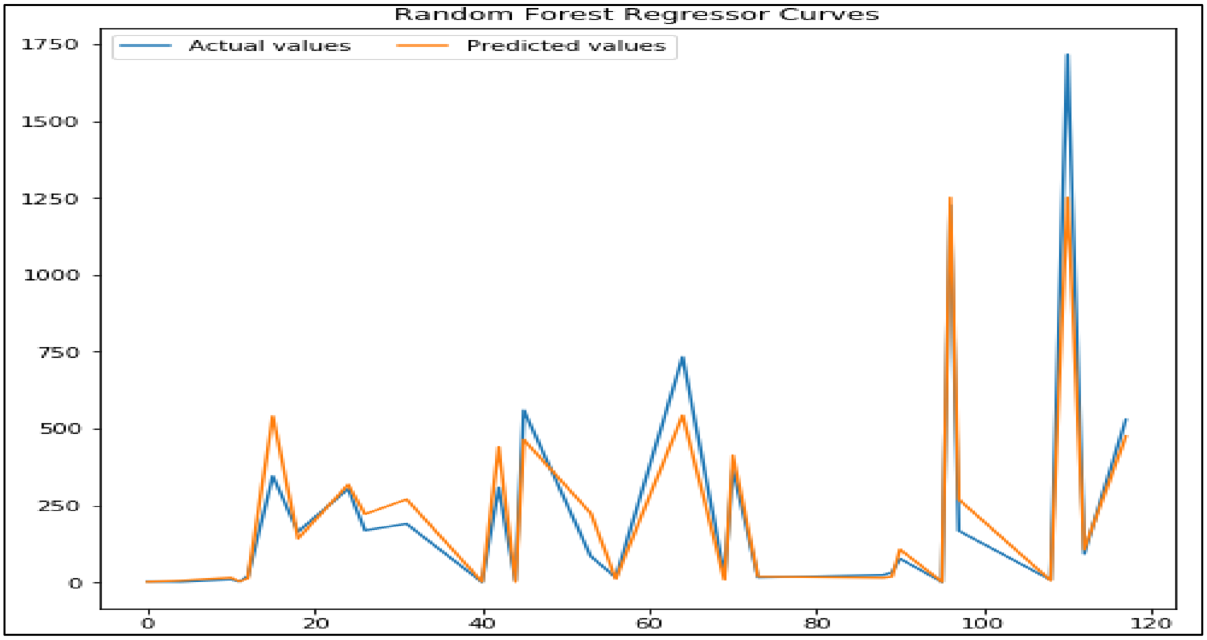


**Figure 10: Random Forest Regressor Prediction Model**

The curve 9 shows that the Random Forest Regressor Prediction Model fits with the Rwandan GBV dataset, which therefore can lead to high and best performance. The following are the predicted values of the row data which have been randomly picked from the entire dataset.

**Table 5: Image Table of RFR Model predicted**

| Random Row Data | 0.000000 | 4.000000 | 10.00000 | 11.000000 | 12.000000 | 15.000000 | 18.000000 | 24.000000 | 26.000000 | 31.000000 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual Values | 3.000000 | 2.000000 | 9.00000 | 3.000000 | 19.000000 | 345.000000 | 164.000000 | 304.000000 | 169.000000 | 190.000000 |
| Predicted Values | 1.962477 | 5.941193 | 14.99317 | 3.145089 | 12.757016 | 540.108578 | 142.087654 | 317.591534 | 221.510709 | 269.681829 |

**Source:** Python RFR prediction results

**Comparative study of the Fitting Models basing on Metrics**

**Evaluation Metrics table (R^2 score, MSE AND RMSE**

Once a machine learning model/Algorithm has been trained, machine learning metrics allow a researcher to measure its performance. In this dissertation, Regression metrics like MSE (Mean Square Error), RMSE (Root Mean Square Error), R^Score and MAE (Mean Absolute Error) have been used to assess the performance of models.

**Table 6: Evaluation of metrics on all used models**

| Method | Training | | | | Testing | | | |
|---|---|---|---|---|---|---|---|---|
| | R 2 score | MSE | RMSE | MAE | R 2 Score | MSE | RMSE | MAE |
| LINEAR REGRESSION | 0.62 | 39865.45 | 199.66 | 144.82 | 0.62 | 30569.13 | 174.84 | 128.84 |
| LASSO | 0.61 | 41420.53 | 203.52 | 140.53 | 0.62 | 30495.20 | 174.62 | 118.84 |
| GRADIENT BOOSTING REGRESSOR | 0.74 | 27262.82 | 165.11 | 118.02 | 0.77 | 18642.99 | 136.53 | 98.38 |
| RANDOM FOREST REGRESSOR | **0.92** | **7513.05** | **86.67** | **44.24** | **0.94** | **4413.85** | **66.43** | **44.44** |

**Source:** Results of metrics from Python

The first algorithm which is Linear Regressor has shown the low value of R^2 score which was 0.62 or 62% and it means that as it does not have a good score, it is not the best algorithm for the current dataset. Apart from the low value of R^2 score, it has counted a big number of errors as calculated by MSE and RMSE.

Lasso algorithm as well has been assessed by R^2 score, and showed that its performance cannot be relied on as it is of low value with 0.62 or 62%. MSE and RSME values are at a big number so that the algorithm with a significant number of errors cannot be trusted as the best for prediction. Moreover, gradient boosting regressor has been used for prediction and the metrics assessment revealed that R^2 score is still at a number which cannot be taken as a good value by having 0.74 or 74% as R^2 score value, in addition of a big value of MSE and RMSE.

Lastly, the assessment of Random Forest Regressor shows different results from other algorithms, where R^2 score scored 94% or 0.94 which is a significant value scoring. Values obtained from other two metrics MSE and RMSE are at a low number comparing to other algorithms. Therefore, basing on evaluation metrics values the Random Forest Regressor has been considered as the best fitting algorithm to forecast the current GBV dataset

**Evaluation Metrics on curves presentation (R^2 score, MSE AND RMSE)**

1. **R^2 Score curve**

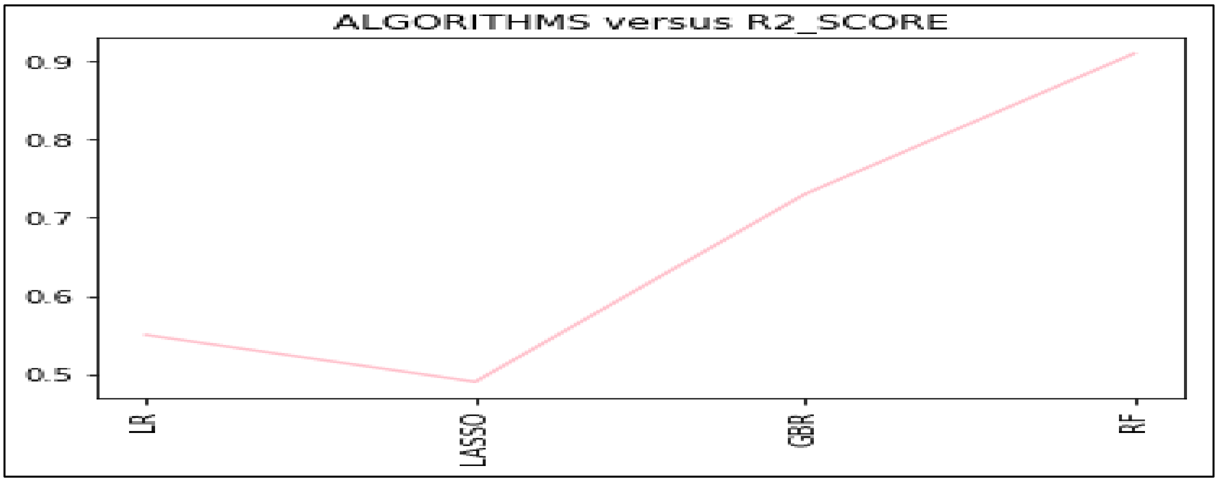


**Figure 11: Algorithms assessment by R^2 score**

**Mean Square Error curve**

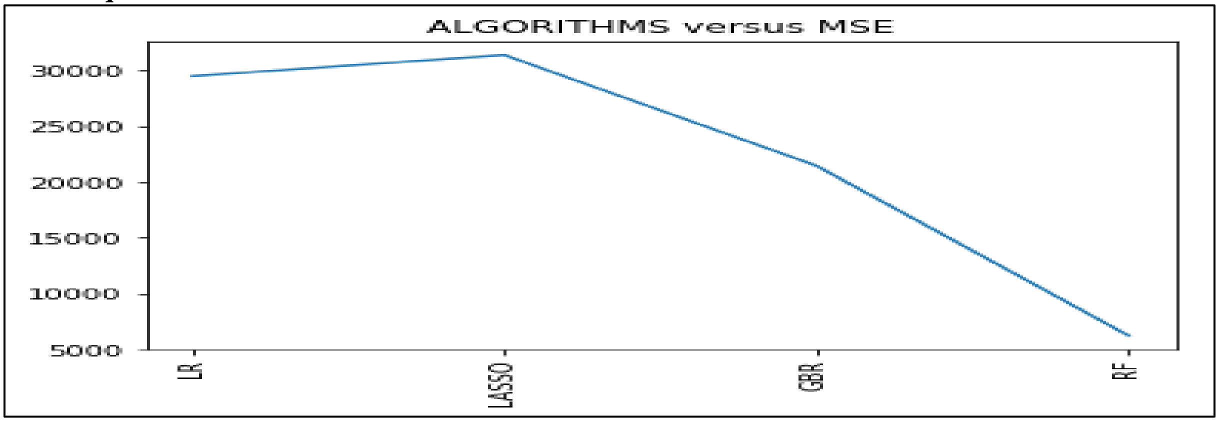


**Figure 12: Algorithms assessment by MSE**

2. **Root Mean Square Error curve**

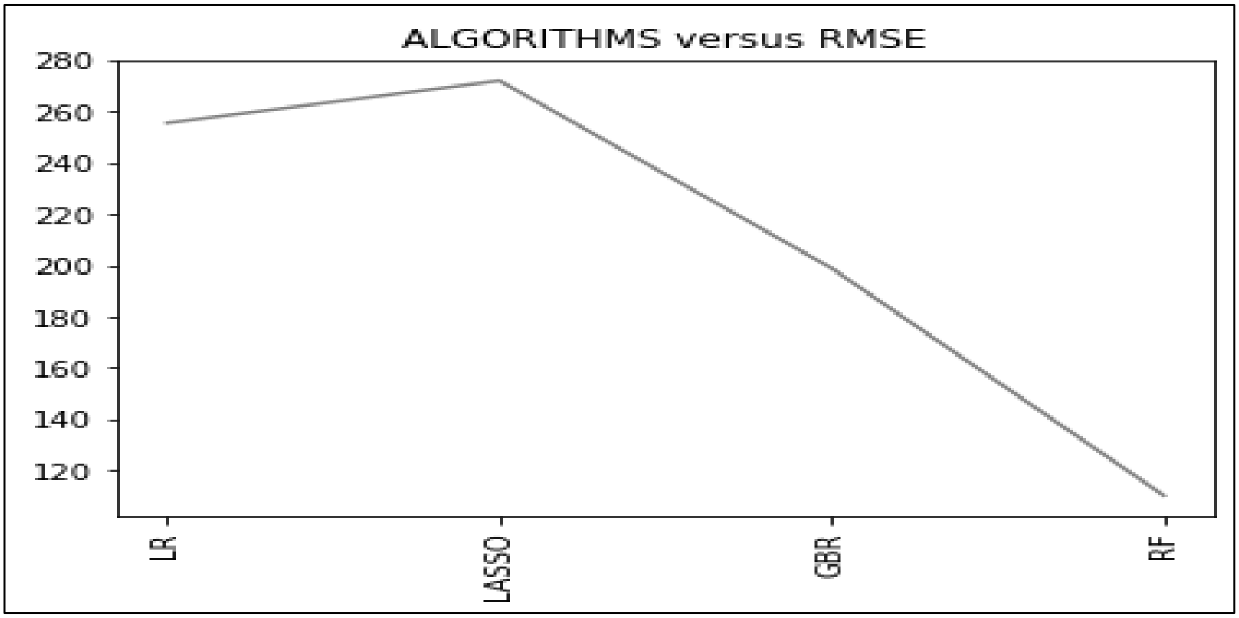


**Figure 13: Algorithms assessment by MSE**

### 3. Mean Absolute Error curve

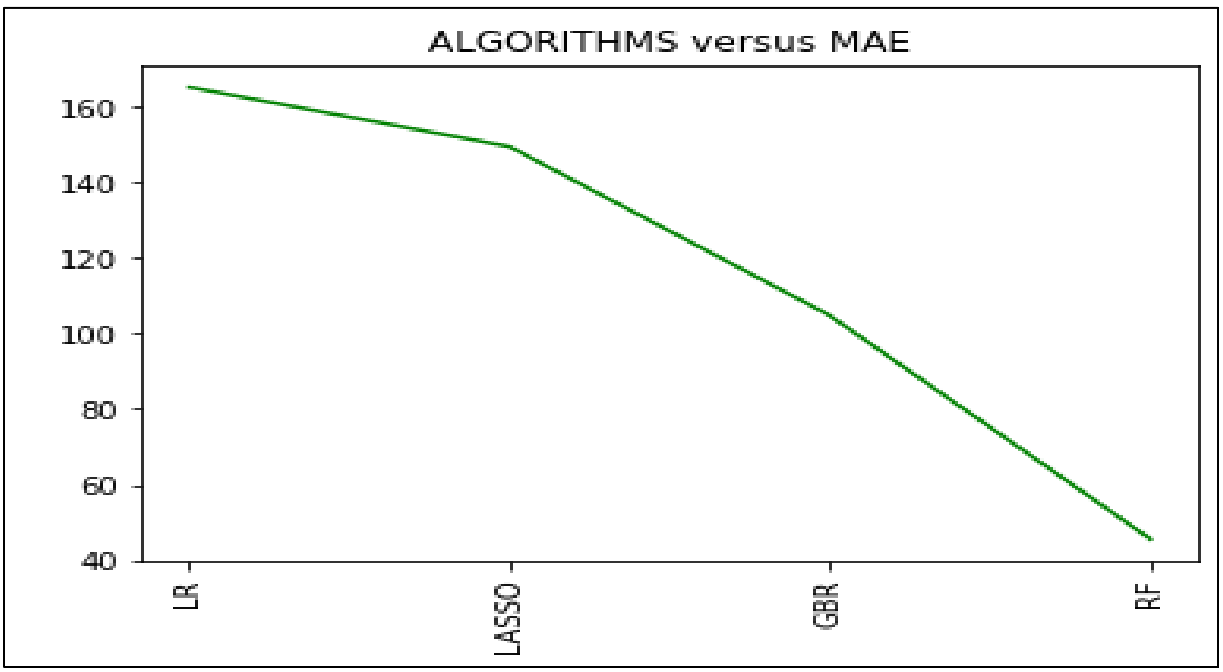


**Figure 14: Algorithms assessment by MAE**

As it is illustrated by above four figures, the R^2 score curve is toping up at Random Forest algorithm, and MSE and RMSE and also at MAE curves show the Random Forest Algorithm at the bottom of the graph, meaning that the value of errors at Random Forest algorithm is not significant. Therefore, Comparative study of the Fitting Models basing on Metrics conclude that Random Forest Algorithm is the best fitting among others, then it is the one to use in GBV Rwandan dataset.

**Findings on the associated factors of Gender-Based Violence in Rwanda.**

As literally stated, Gender-based violence is violence that is directed against a person on the basis of their gender or sex, including acts that inflict physical, mental or sexual harm or suffering, threats of such acts, coercion and other deprivations of liberty. It includes physical, sexual and psychological violence perpetrated or condoned within the family, the general community or by the state and its institutions. During this study at Rwanda, respondents confirmed that such violence can take many different forms.

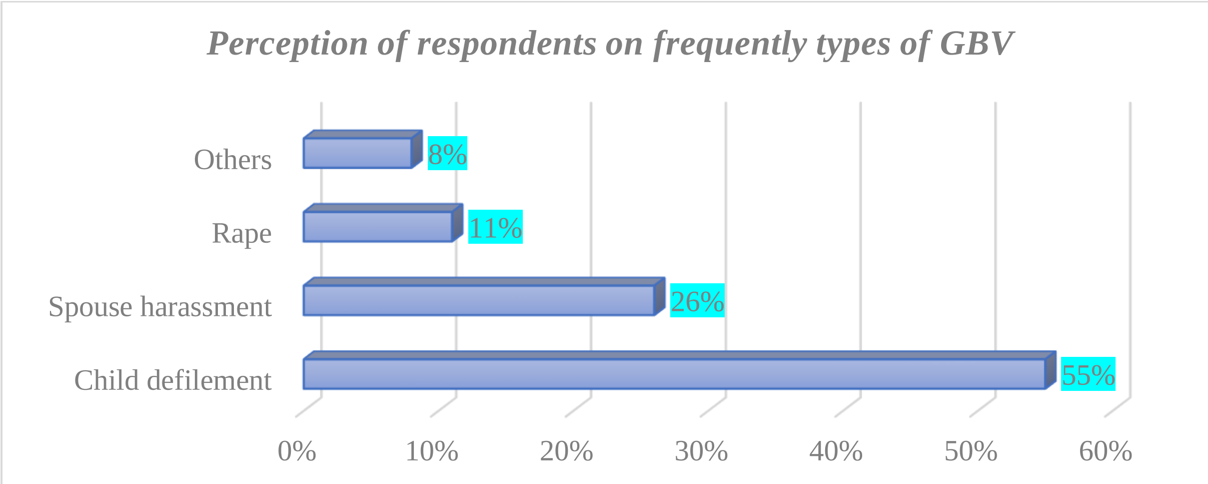


**Figure 15: Perceptions of respondents on categories of gender-based violence experienced in Rwanda;**
**Source:** *Primary data, (2022)*

Findings indicated that in Rwanda reported cases for GBV; the use of Child defilement was confirmed by 55% of respondents; 26% respondents argued in Spouse harassment; 11% of respondents confirmed Rape cases, and others like sexual exploitation, female genital mutilation, the forced marriage and early marriages cases of GBV were confirmed by on rate of 8% of respondents in Isange One Stop Center Kacyiru, and RWAMUREC in Rwanda.

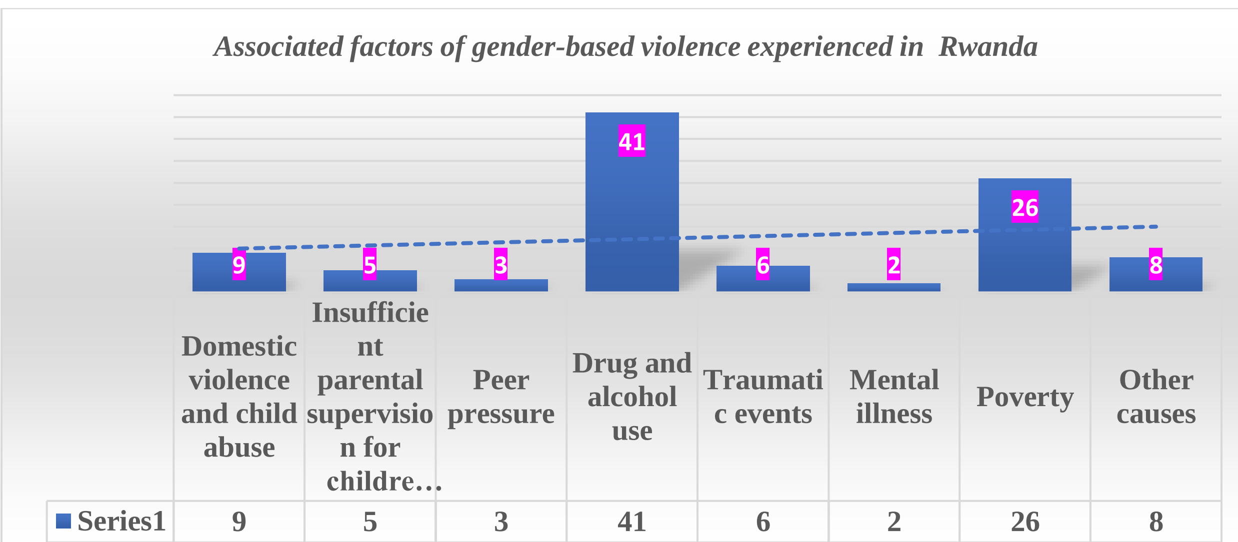


**Figure 16: Perceptions of respondents on causes or associated factors of gender-based violence experienced in Rwanda**
**Source:** *Primary data, (2022)*

Finding in figure 15 present the perceptions of respondents on causes or associated factors of gender-based violence experienced in Rwanda. The results revealed that domestic violence and child abuse are among the associated factors of GBV as confirmed by 9.0% of the respondents; insufficient parental supervision for children victims of GBV, confirmed on rate of 5.0%; peer pressure was stated on rate of 3.0%; drug and alcohol use is highly confirmed by the majority of 41.0% as main associated factors of GBV in Rwanda; traumatic events is as associated factor of GBV as confirmed by 6.0%; mental illness was confirmed on rate of 2.0%; poverty especially in households are stated as among the associated factors of GBV on rate of 26.0%; while other causes or associated factors unstated in this survey, confirmed on rate of 8.0% in all respondents participated in this survey at Rwanda.

## CONCLUSION

As it is illustrated by above three figures, the R^2 score curve is toping up at Random Forest algorithm, and MSE and RMSE curves show the Random Forest Algorithm at the bottom of the graph, meaning that the value of errors at Random Forest algorithm is not significant. Lasso Regression, Gradient Boosting, and applied these algorithms on the dataset. From this study, on our dataset, Random Forest regression performed the best score. On the contrary LASSO Regression performed the least. Therefore, Random Forest Regressor has been judged as the best forecasting algorithm on GBV dataset in Rwanda.

## RECOMMENDATIONS

The pre-processing of data revealed that at the recording stage, they should include the specific date of the crime so that it can be treated it time series. Also recording the scene of crime would be better if the province, district, sectors and villages are recorded separately.

According to factors associated with GBV in Rwanda, where the consumption of drugs comes at the first rank with a high percentage, it is advised to reinforce community policing and to empower the culture of reporting GBV crimes at the real time so that other organs in charge of it should also intervene at the real time for victim.

## REFERENCES


[1.] Abrahams, N., et al. (2019). *Intimate Partner Femicide in South Africa in 1999 and 2009.* PLoS medicine.

[2.] Addisu Shunu Beyene , Catherine Chojenta et al. (2021). *Gender-Based Violence Among Female Senior Secondary School Students in Eastern Ethiopia.*

[3.] Alemu Basazin & Tadesse. (2021). Prevalence and associated factors of gender based violence among Baso high school female students, 2020. *Reproductive Health*.

[4.] Alemu Basazin Mingude & Tadesse Mamo Dejene . (2021). *Prevalence and associated factors of gender based violence among Baso high school female students, 2020.*

[5.] Beyene, Chojenta C, Roba, et al. (2019). *Gender-based violence among female youths in educational institutions of Sub-Saharan Africa: a systematic review and meta-analysis.* Syst Rev. 2019;8(1):1–14.

[6.] Bloom, Shelah. (2018). *"Violence Against Women and Girls: A Compendium of Monitoring and Evaluation Indicators.".* Carolina Population Center, MEASURE Evaluation, Chapel Hill, North Carolina.: https://www.measureevaluation.org/resources/publications/ms-08-30.

[7.] Decker et al. (2017). Gender-based violence against adolescent and young adult women in low- and middle-income countries. *The Journal of Adolescent Health*, 56(2): p. 188-96.

[8.] Greg Van Houdt. (2020). *A Review on the Long Short-Term Memory Model.* DOI:10.1007/s10462-020-09838-1.

[9.] Habtamu Gebrehana Belay. (2021). *Magnitude of Gender-Based Violence and Its Associated Factors among Female Night Students in Bahir Dar City, Amhara Region, Ethiopia.*

[10.] Ifeoma Bridget Udigwe et al. . (2021). *Prevalence and Types of Gender-Based Violence among Secondary School Students During Covid-19 Pandemic in Anambra State, South-East Nigeria.*

[11.] Kvinna till Kvinna Foundation . (2021). *Gender-based violence: Sweden.* Johanneshov: https://kvinnatillkvinna.org/about-us/what-we-do/gbv/.

[12.] Meghan Ott. (2021). *Gender-based violence (GBV) includes physical, sexual, verbal, emotional, and psychological abuse, threats, coercion, and economic or educational deprivation.* https://www.womenforwomen.org/blogs/series-what-does-mean-gender-based-violence#:~:text=Gender%2Dbased%20violence%20(GBV)%20is%20violence%20that

%20is%20directed,in%20public%20or%20private%20life.

[13.] MIGEPROF. (2021). *Prevelence of GBV of Rwanda between 2019-2020.* https://www.migeprof.gov.rw/news-detail/government-ups-efforts-to-tackle-gender-based-violence#:~:text=The%20Rwanda%20Investigation%20Bureau%20statistics,in%20the%20previous%20fiscal%20year.

[14.] Mohammed Kemal. (2019). *Data Visualization: Methods, Types, Benefits, and Checklist.* Building an Enterprise Data Warehouse.

[15.] Nabukeera M. (2020). *Prevention and response to gender-based violence (GBV) during novel Covid-19 lock-down in Uganda.* Kampala: J Adult Protection. 2020;23:116.

[16.] Nazri M., and Walid Hasen Atomia. (2013). *The effect of Data Pre-Processing on Optimized Training of Artificial Neural Networks.* Parit Raja, Batu Pahat, Johor, Malaysia.: The 4th International Conference of Electrical Engineering and Informatics (ICEEI 2013).

[17.] Olufunmilayo Fawole, Olubunmi D Balogun et al. (2018). *Experience of gender-based violence to students in public and private secondary schools in Ilorin, Nigeria.*

[18.] Palermo, and Peterman A. (2016). *Tip of the iceberg: reporting and gender-based violence in developing countries.* Am J Epidemiol. 2014;179:602–12.

[19.] Richters, A., Rutayisire,et al. (2018). *Care as a turning point in sociotherapy: remaking the moral world in post-genocide Rwanda.* . In: Medische anthropologie 22(1) .

[20.] Rurangirwa, Mogren, Ntaganira, and Gunilla Krantz. (2017). Intimate partner violence among pregnant women in Rwanda, its associated risk factors and relationship to ANC services attendance: a population-based study.

[21.] Stephen Ssenkaaba. (2018). *Uganda: Violence against women unabated despite laws and policies despite laws and policies.* https://www.un.org/africarenewal/news/uganda-violence-against-women-unabated-despite-laws-and-policies#omen%20is%20on,cases)%20between%202015%20and%202016.

[22.] UNFPA State. (2016). *Of World Population 2016: By Choice, Not by Chance.* New York, NY: United Nations Population Fund; 2016.

[23.] Vinay, Kumar, Neelam; Malik, et al. (2022). Socio demographic determinants of violence among school-going adolescent girls in a rural area of North India.