

Original article**A robust hybrid methodology between applied linear regression model (alm) and multilayer perceptron (mlp)**

Mohamad Nasarudin Bin Adnan¹, Wan Muhamad Amir W Ahmad^{a2}, Nuzlinda Abdul Rahman³, Farah Muna Mohamad Ghazali⁴, Nor Azlida Aleng⁵, Zainab Mat Yudin @ Badrin⁶, Mohammad Khursheed Alam⁷, Nor Farid Mohd Noor⁸

Abstract

Background: The goal of this study is to illustrate an optimum variable selection method using established Multiple Linear Regression (MLR) models and to validate the variable using Multilayer Perceptron Neural Network (MLP) models. Initially, all selected variables will be passed through the bootstrap methodology, and they were screened for significant relationships. **Objective:** The goal of this work is to analyze and construct a model for the factor linked with total crime cases by combining an Applied Linear Regression Model (ALRM) and a Multilayer Perceptron (MLP). **Material and Methods:** Around 200 data was simulated to build the methodology. Advanced computational statistical modeling methodologies were used to evaluate data descriptions of several variables in this retrospective study, including the total victim, gender, age, marital status, social class, adult in the household, children in household, burglary's victim, sexual's victim, victim's report, and household location. The case study was developed and implemented using the R-Studio program and syntax. **Results:** The statistical method demonstrated that regression modeling surpasses R-squared and mean square error test in most situations. Researchers observed that when data is divided into two datasets for training and testing, the hybrid model approach performs significantly better at predicting the experiment's outcome. When it came time to determine variable validity, the well-established bootstrap-integrated MLR approach was applied. Ten characteristics are taken into consideration in this case: Gender (β : -0.4369700; $p < 0.25$), age (β : -0.0086757; $p < 0.25$), marital status (β : 0.2646097; $p < 0.25$), social class (β : 0.0602540; $p < 0.25$), adult in household (β : -0.0211293; $p > 0.25$), children in household (β : -0.0025346; $p > 0.25$), burglary's victim (β : 1.3473593; $p < 0.25$), sexual's victim (β : 1.0382444; $p < 0.25$), victim's report (β : -0.3176104; $p < 0.25$), and location of household (β : -0.1355046; $p < 0.25$). There is a 0.07745823 MSE for the linear model in this scenario. **Conclusion:** The neural network's Predicted Mean Square Error (PMSE) was used to assess MLP's performance (MSE-forecasts the Network). PMSE is used to determine how far our projections are from the actual data, and the lowest MSE from the MLP indicates the best achievement. The R syntax for MLR and MLP is also included in this research article. As a result, the study's conclusion establishes the superiority of the hybrid model technique.

Keywords: Multiple Linear Regression; Multilayer perceptron

Bangladesh Journal of Medical Science Vol. 22 No. 01 January'23 Page : 38-46
DOI: <https://doi.org/10.3329/bjms.v22i1.61850>

1. Mohamad Nasarudin Bin Adnan
2. Wan Muhamad Amir W Ahmad
3. School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia. wmamir@usm.my; nasarudinadnan@student.usm.my; munaghazali@yahoo.com; drzainab@usm.my
4. Nuzlinda Abdul Rahman, Farah Muna Mohamad Ghazali, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia. nuzlinda@usm.my
5. Nor Azlida Aleng^c, Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia. azlida_aleng@umt.edu.my
6. Zainab Mat Yudin @ Badrin, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia.
7. Mohammad Khursheed Alam, Orthodontic Unit, College of Dentistry, Jouf University, Sakaka, Saudi Arabia (mkalam@ju.edu.sa; dralam@gmail.com) & Department of Public Health, Faculty of Allied Health Sciences, Daffodil International University. Dhaka, Bangladesh.
8. Nor Farid Mohd Noor, Faculty of Medicine, Universiti Sultan Zainal Abidin (Uni SZA), Medical Campus, Jalan Sultan Mahmud, 20400 Kuala Terengganu, Terengganu, Malaysia. faridmnoor@unisza.edu.my

Correspondence: Mohammad Khursheed Alam, Orthodontic Unit, College of Dentistry, Jouf University, Sakaka, Saudi Arabia (mkalam@ju.edu.sa; dralam@gmail.com) & Department of Public Health, Faculty of Allied Health Sciences, Daffodil International University. Dhaka, Bangladesh.

Introduction

Regression modelling is based on a combination of two phenomena that can be represented by equations involving dependent and independent variables, one of which is known as Applied Linear Regression Model (ALRM)^{1,7}. It demonstrates how the dependent variable responds to a changes in the independent variables, as well as which independent variable is related to the dependent variable, regardless of whether the relationship is linear or nonlinear⁸.

The multiple linear regression (MLR) is very famous technique and extensively used in many fields especially in medical field³. It is used for modelling the relationship between two or more independent variables and a continuous dependent variable. It's used when there are two or more independent variables and a single dependent variable. MLP $Y = \hat{a}_0 + \hat{a}_1 \div_1 + \hat{a}_2 \div_2 + \dots + \hat{a}_k \div_k$ is an extension of simple linear regression which used in this study to get a better results⁶. Cohen et al. (2003) used the MLR in a study to determine the factors that influence nurses' attitudes toward their patients' families. The regression analysis was based on the robust standard errors as the assumption of normality was violated. The results show that there is no problem were found, and the variables had variance inflation factor (VIF) values ranging between 1.00 to 1.36⁵.

Neural network are the modelling tools for neurophysiology and artificial intelligence¹⁴. However, they are also apply as statistical models in medical field describe and make inferences from the data, and to increase the range of epidemiological problems mostly using multilayer perceptron (MLP)⁴. Many epidemiological studies provide insufficient information about the covariates' statistical properties, such as normality of distribution patterns, existence of colinearity, and unknown of link function between variables^{1,9,14}. Thus, the MLP appears to be a solution as three layer perceptron networks are theoretically universal approximators¹². Moreover, some studies claim they can match or exceed the classical statistical methods in terms of goodness of fit, estimation, and prediction^{3,13}.

The purpose of this study is to analyse and construct the Applied Linear Regression Model (ALRM) and a Multilayer Perceptron (MLP) model for the factor linked with the total crime cases

Materials and Methods

Data Simulation

This study examined simulation data from a trial that enrolled 200 observers. Table 1 lists the data descriptions for the research variables.

Table 1. Data Description of the research variables

Code	Variables	Descriptions
Y	Total victim	Number of times respondent victimized in year
X ₁	Gender	Gender 1 = Male 2 = Female
X ₂	Age	Age in years
X ₃	Marital	Marital status 1 = Married, 2 = Divorced, 3 = Separated 4 = Other
X ₄	Socclass	Social class 1 = Daily base working, 2 = Middle class 3 = Upper middle class, 4 = Not working
X ₅	Adult	Number of adults in the household
X ₆	Children	Number of children in the household
X ₇	Burglary	Number of times victimized by burglary
X ₈	Sexual	Number of times victimized by the sexual offense.
X ₉	Report	Number victimizations reported to the police
X ₁₀	Location	Location of household 1 = Urban , 2 = Sub Urban , 3 = Rural

Study Design

This study employs a simulation data approach and advanced computational statistical modelling methodologies to build a multilayer perceptron with multiple linear regression methodology. This developed methodology was based on the testing and training dataset, MSE-predicted, and the mean absolute deviance (MAD) accuracy value.

Computational Biometry Modeling

Multiple linear regression was used to analyze the simulation data. In the R-Studio software, the R-syntax was developed with a successful combination model that used bootstrap, multiple linear regression, and multilayer perceptron approaches. These advanced techniques led to the superiority of the hybrid model technique. The data is divided into two categories: training and testing. Modeling will be done with the training data, while validation will be done with the testing data. This section investigates the relationship between the total number of cases and the selected explanatory variables by fitting a set of linear regression models.

$$\tilde{O} = \hat{\alpha}_0 + \hat{\alpha}_1(\text{Gender}) + \hat{\alpha}_2(\text{Age}) + \hat{\alpha}_3(\text{Marital}) + \hat{\alpha}_4(\text{Socclass}) + \hat{\alpha}_5(\text{Adult}) + \hat{\alpha}_6(\text{Child}) + \hat{\alpha}_7(\text{Burglary}) + \hat{\alpha}_8(\text{Sexual}) + \hat{\alpha}_9(\text{Report}) + \hat{\alpha}_{10}(\text{Location}) \dots \dots \dots (1)$$

Where β_0, \dots, β_0 are parameters, and Y is the number of times respondents were victimized in a year.

Bootstrap

Bootstrap calculates sample statistics by starting with a random sample of the population. A pseudo-population is created by using a large number of substitution samples that are replicated several times after the original samples have been replicated. In the case of replacement sampling, random sampling yields samples that differ from the original sample. The bootstrap computes statistics for each sample drawn with replacement and is used to draw replacement samples. (Efron, 1982; Efron and Tibshirani, 1994).

Table 2 displays the model's outcome. The linear regression must be fit, and the R software is used. The complete step-by-step procedure is as follows:

A diagram of this process is shown in Figure 1. One of the study's strengths is that it looks into a model that takes clinically relevant variables into account. Therefore, a bootstrapping method will be developed after the data preparation. The bootstrap method generates a sample of the same size as the original sample, but each observation is repeated several times, and others are discarded (Efron, 1982; Efron and Tibshirani, 1994).

The syntax in R

#!/Dataset for Biometry: Forensic Modeling Study #

```
Input =("
gender age marital socclass adult child total burglary
sexual report location
1 35 1 1 5 0 3 2 1 1 1
1 26 1 1 3 2 3 2 1 1 1
1 29 2 1 4 1 0 0 0 0 1
1 19 2 2 1 2 0 0 0 0 3
1 27 1 2 1 0 1 0 1 0 2
1 43 1 1 2 0 0 0 0 0 1
2 61 2 1 3 0 3 3 0 0 1
2 43 1 2 2 1 2 1 1 1 1
2 22 2 2 1 3 0 0 0 0 1
2 56 3 1 2 1 0 0 0 0 2
2 43 2 3 1 2 1 1 0 1 1
2 43 2 3 1 2 1 1 0 1 1
:::
2 43 2 3 1 2 1 1 0 1 1
2 43 2 3 1 2 1 1 0 1 1
1 55 1 2 2 1 2 2 0 1 1
2 32 2 1 3 2 4 3 1 3 1
1 40 1 1 2 0 6 4 1 2 1
```

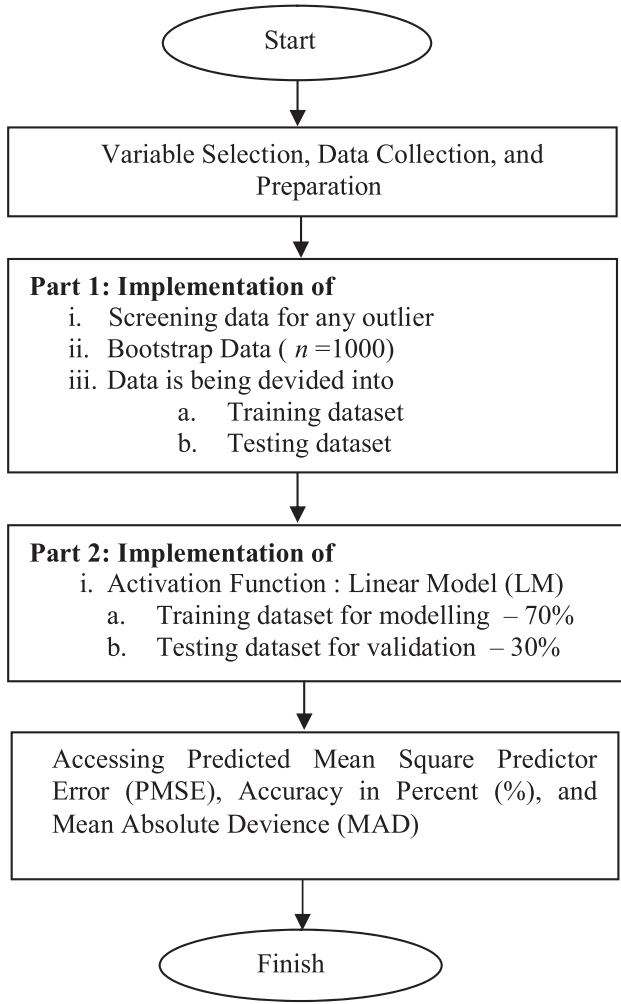


Figure 1. Flowchart of the proposed statistical-linear modeling

```

1 32 3 1 4 0 3 2 0 1 1
1 23 2 2 5 0 2 0 2 1 2
1 25 4 2 2 0 5 2 1 0 2
2 34 3 3 1 0 2 1 1 0 2
2 23 2 2 1 0 0 0 0 0 1
2 22 2 2 1 3 0 0 0 0 1
2 56 3 1 2 1 0 0 0 0 2
2 43 2 3 1 2 1 1 0 1 1
2 43 2 3 1 2 1 1 0 1 1
1 55 1 2 2 1 2 2 0 1 1
“)
data = read.table(textConnection(Input),header=TRUE)
#/Performing Bootstrap for 1000 : Case Resampling Procedure /#
mydata <- rbind.data.frame(data, stringsAsFactors = FALSE)
iboot <- sample(1:nrow(mydata),size=1000, replace = TRUE)
bootdata <- mydata[iboot,]
#/Install the Neuralnet Package /#
if(!require(neuralnet)) {install.packages(“neuralnet”)}
library(“neuralnet”)
#/Checking for the Missing Values /#
apply(bootdata, 2, function(x) sum(is.na(x)))
#/Scaling the Data for Normalization /#
#/Method (Usually Called Feature Scaling) to Get All the Scaled Data /#
#/In the Range [0,1] /#
max_data <- apply(bootdata, 2, max)
min_data <- apply(bootdata, 2, min)
data_scaled <- scale(bootdata, center = min_data, scale = max_data - min_data)
#/Randomly Split the Data into 70:30 /#
#/70 Percent of the Data at Our Disposal to Train the Network /#
#/30 Percent to Test the Network /#
index = sample(1:nrow(bootdata),round(0.70*nrow(bootdata)))
train_data <- as.data.frame(data_scaled[index,])

test_data <- as.data.frame(data_scaled[-index,])
#####
#####

#/Build the Network /#
#/There are 3 Hidden Layers Have 3 and 2 Neurons Respectfully /#
#Input = 10 /#
#Output = 1 /#

n = names(bootdata)
f = as.formula(paste(“total ~”, paste(n[!n %in% “total”], collapse = “ + “)))
nn=neuralnet(f,data=train_data,hidden=c(4,3),linear.output=T)
plot(nn)
options(warn=-1)

#/30 Percent of the Available Data to do this: Using Only the First 2 Columns Representing the Input Variables /#
#/of The Network and 1 is The Output for NN /#
predicted <- compute(nn,test_data[,1:10])

#/Use the Mean Squared Error NN (MSE-forecasts the network) as a Measure of How Far Away Our Predictions Are From The Real Data /#
MSE.net <- sum((test_data$total-predicted$net.result)^2)/nrow(test_data)
MSE.net
#####
#####

Model m(total~gender+age+marital+socclass+adult+child+burglary+sexual+report+location,
data=bootdata) # build the model
summary(Model)
data$Predictedtotal <- predict(Model,data)
distPred <- predict(Model, data)

```

```

preds <- predict(Model, data)
modelEval <- cbind(data$total, preds)
colnames(modelEval) <- c('Actual', 'Predicted')
modelEval <- as.data.frame(modelEval)
print (modelEval)

test <- data[,-index,]
predict_lm <- predict(Model,test)
MSE.lm <- sum((predict_lm - test$total)^2)/
nrow(test)
MSE.lm

```

```

#!/Printing the Value of MSE for Linear Model and Neural Network/
print(paste(MSE.lm,MSE.net))

```

Finished

Results

This research aims to look into the performance of a Multilayer Perceptron based on the linear activation

function: it takes into account both the training and testing datasets. The best multiple linear regression model was found by selecting variables with the lowest Predicted Mean Square Error as determined by the multiple linear regression algorithm.

Method of MLP-based regression without bootstrapping

Table 2 displays the results of multiple linear regression using a training dataset, where the total victim is a dependent variable that serves as the study's outcome. The linear model's square error net (MSE.net) prediction is 0.01318659 in this case, while the predictive Mean Square Error (MSE.lm) is 0.07239172. These values show how far apart the predicted and actual values are. Also, demonstrate how dispersed our predictive data is and how good the prediction performance of the obtained model is. Remarkably similar data will have a small spread after, and it speaks to the accuracy and reliability of our forecast data. The train-to-test split is 70:30, which means that 70% of the data is used to train the network, and 30% is used to test the network. The results of the multiple regression analysis are presented in Table 2. Below is the model.

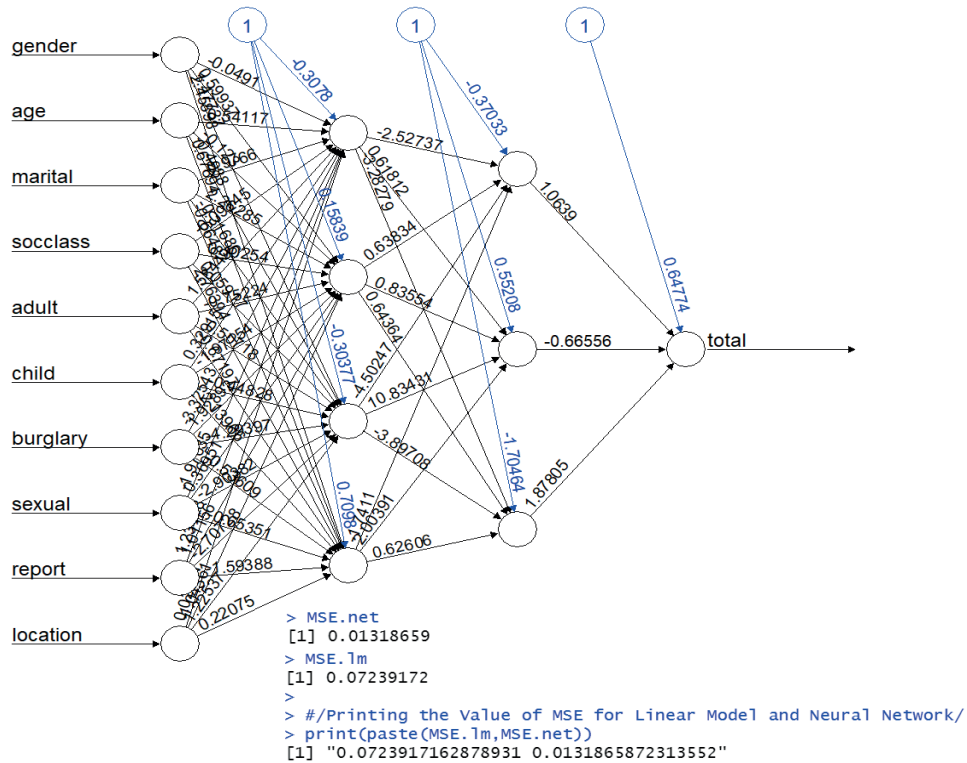


Figure 1.1 The architecture of the best (MLP) model with ten input variables, two hidden layers, and one output node (Proposed Model without bootstrap)

Table 2. Result of Multiple Logistic Regression with combining the bootstrap method training and testing dataset

Variable	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	1.013923	0.116641	8.693	< 2e-16*
Gender	-0.607740	0.042369	-14.344	< 2e-16*
Age	-0.004515	0.001257	-3.591	0.000363*
Marital	0.349041	0.020738	16.831	< 2e-16*
Socclass	0.013473	0.018644	0.723	0.470231
Adult	-0.080656	0.017202	-4.689	3.57e-06*
Child	0.005867	0.016085	0.365	0.715450
Burglary	1.274513	0.021875	58.263	< 2e-16*
Sexual	1.136620	0.040622	27.981	< 2e-16*
Report	-0.266752	0.031813	-8.385	5.46e-16*
Location	-0.301678	0.046150	-6.537	1.58e-10*

Multiple Linear Regression was applied; *Significant at the level of the 0.05 R^2 : 98.16%

The variables in this section were validated using the established bootstrap method, which is embedded in the developed syntax in integrated multiple linear regression. Ten variables were chosen for analysis in this case: gender, age, marital status, social class,

adult in the household, children in household, burglary victim, sexual victim, victim's report, and household location. All of the variables were found to have a significant impact on the overall victim. Table 3 summarizes the detailed output.

Method of MLP-based regression with bootstrapping

In this section, the bootstrapping method was applied. Results of multiple linear regression using a training set are shown in Table 3. dataset, where the total victim is a dependent variable that serves as the study's outcome. The linear model's square error net (MSE.net) prediction is 0.006816 in this case, while the predictive Mean Square Error (MSE.lm) is 0.077458. These numbers represent the difference between the predicted and actual values. It also demonstrates how dispersed our predictive data is and how good the prediction performance of the obtained model is. This small spread indicates our forecast data is accurate and reliable. Training and testing data are split 70:30, meaning 70% of the data is used to train the network and 30% for testing. Table 3 shows the results of multiple regression analysis. Below is the model

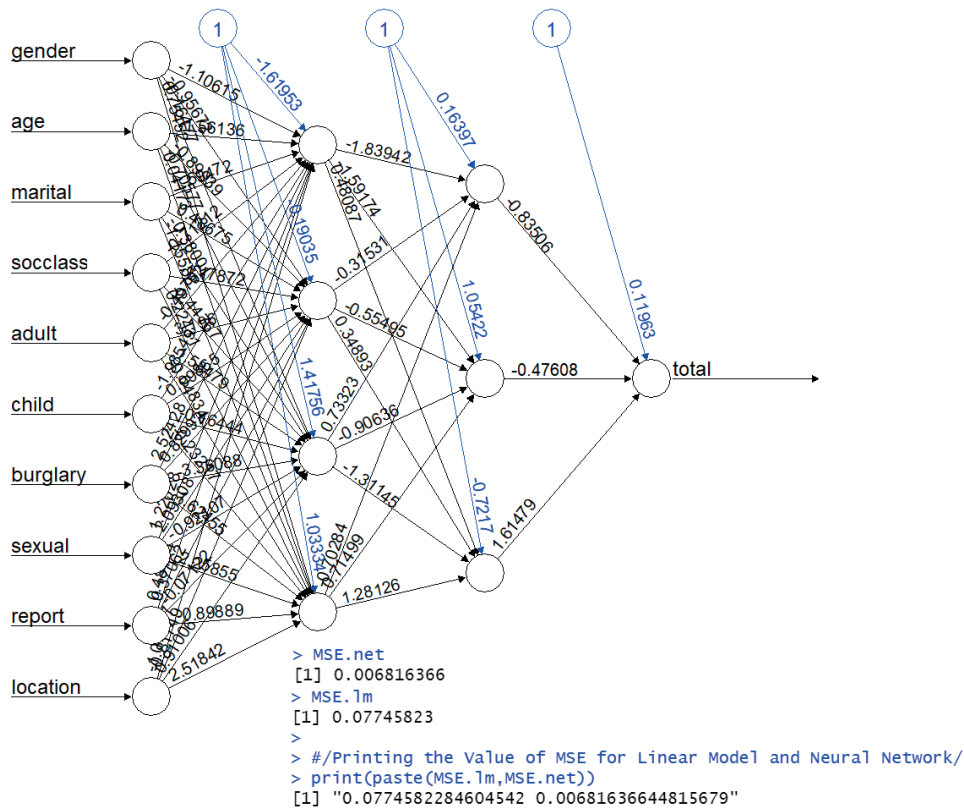


Figure 1.2 The architecture of the best (MLP.) model with ten input variables, one hidden layer, and one output node (Obtained model with bootstrap)

Table 3. Result of Multiple Logistic Regression with combining the bootstrap method training and testing dataset

Variable	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	0.6473736	0.0808129	8.011	3.19e-15*
Gender	-0.4369700	0.0278577	-15.686	< 2e-16*
Age	-0.0086757	0.0007827	-11.084	< 2e-16*
Marital	0.2646097	0.0128154	20.648	< 2e-16*
Socclass	0.0602540	0.0122312	4.926	9.82e-07*
Adult	-0.0211293	0.0107434	-1.967	0.0495*
Child	-0.0025346	0.0116126	-0.218	0.8273
Burglary	1.3473593	0.0143353	93.989	< 2e-16*
Sexual	1.0382444	0.0275570	37.676	< 2e-16*
Report	-0.3176104	0.0216068	-14.700	< 2e-16*
Location	-0.1355046	0.0325692	-4.161	3.45e-05*

Multiple Linear Regression was applied; *Significant at the level of 0.25 R^2 : 98.23%

Model evaluation of the model

The model is evaluated based on the forecast value. Actual and predicted values will be compared to determine accuracy. Testing data set used to evaluate model built from training data set. The distance prediction will be used to compare the predicted and actual data. The R syntax provides a model assessment approach that can be used to evaluate subsequent methods. The “Actual” and “Predicted” values obtained from the proposed methodology are shown in Table 3.

Table 3. The “Actual and “Predicted” values obtained through the proposed methodology

Actual	Predicted
3	3.40581852
3	3.52108942
0	0.32572470
0	0.26257994
1	1.10738271
0	-0.01555183
3	3.67687361
2	1.67319081
0	0.30689568
0	0.28302418
2	2.25521073
1	1.04515624
0	-0.48392287
6	5.80293610
3	2.94395010

Actual	Predicted
2	2.04281198
5	4.59215268
2	2.54651546
0	0.06698559

The difference between the “Actual” and “Predicted” values is minimal. Furthermore result of a paired sample t-test shows that there are no significant differences.

Table 4 Comparison of the “Actual Data” with “Predicted Data”

Paired Samples Test	
Variables	Mean (SD)
Actual	1.78(1.790)
Predicted	1.8610(1.77)
T Statistics (d.f)	1.673(18)
p-value	0.112
Paired Sample Correlation (ρ)	0.984
p-value	0.000*

*significant at 0.05

Independent samples T-Test was applied Assumptions normality is fulfilled

Table 4 displays the “Actual” and “Predicted” values based on the proposed model. According to the findings, the terms “actual” and “predicted” were statistically indistinguishable. This demonstrates the proposed model’s superiority.

Discussion

Using the proposed method, we were able to estimate the predicted value of a dependent variable accurately. An extremely accurate and reliable model was created using the hybrid method that we discovered and harmonized. The findings revealed that nine variables were chosen for analysis in this case: gender, age, marital status, social class, adult in the household, burglary victim, sexual victim, victim’s report, and household location were significant. This paper focuses on developing multiple linear regressions (MLR) methodologies with Multilayer Perceptron (MLP). Developing, testing, and validating a regression model was the study’s primary objective. The selection of variables incorporates expert advice. The bootstrap method uses the initial data set to create a “mega” file at the start of the operation. The bootstrap procedure, on the other hand, generates a large number of replacement files.

The bootstrap method generates and saves statistical samples. As a final point, the bootstrap method iteratively repeats this process, often thousands of times. Then, finally, the data is ready for analysis. This proposed methodology is based on the R syntax algorithm. The first stage in this proposed methodology is to seek the advice of an expert when selecting variables. Then, the bootstrap is applied to the dataset. Training and testing data will be kept separate. The R syntax algorithm connects the application to the concept of method-based methodology. The first step is to select variables with the assistance and guidance of an expert. Following that, the bootstrap procedure will be applied to the data. At this point, 70% of the bootstrap data will be designated as training data and 30% as testing data. For both building and testing the model, we will use data from the training dataset. A successful model will have the lowest mean absolute deviation. The developed syntax was used to compute the difference between actual and predicted values. The study's findings helped the decision-maker get the best possible outcome. The most difficult tasks are selecting appropriate input parameters, preparing data for linear modeling, and standardizing it. This study found that statistical formulations, computation in R syntax, and the multiple linear regression package

resulted in highly successful linear modeling.

Conclusion

This research focuses on the development of new hybrid methods based on bootstrapping and multiple linear regression. The methodology design concept is based on the R syntax to ensure that the researchers fully comprehend the illustration. In this study, the total victim was the dependent variable. Gender, age, marital status, social status, adult in the household, children in the household, burglary's victim, sexual's victim, victim's report, and household location were independent variables. According to the regression theory, the higher the R squared value, the more accurate the model. After bootstrapping method, it can be seen clearly that the hybrid model performance can measure much better than the conventional method. Furthermore, the test's fit (PMSE-.net) can be evaluated by calculating the smallest predictive value from the obtained model. As a result, a hybrid model can help us better understand the utility of the method and its relative contribution to the outcome.

Acknowledgment

This study was funded by the Ministry of Higher Education (MoHE) Fundamental Research Grant Scheme (FRGS) 203.PPSG.6711666, and the authors would like to thank Universiti Sains Malaysia (USM).-

References

1. Ahmad, W.M.A.W., Awang Nawati, M.A., Aleng, N.A., Shafiq, M. . An alternative method for multiple linear model regression modeling, a technical combining of robust, bootstrap and fuzzy approach. *Journal of Modern Applied Statistical Methods*.2016;**15**(2):
2. Ahmad, W.M.A.W., Shafiq, M., Mokhtar, K., Aleng, N.A., & Rahim, H.A. (c). Simple response surface methodology using RSREG (SAS). *Journal of Modern Applied Statistical Methods*.2016;**15**(1); 855-867.
3. Ahmad, W.M.A.W., N. A. (). Modified Nonlinear Model for Exponential Growth Method and Its Application in Biostatistics Using SAS. *International Journal of Multidisciplinary Research and Modern Education*.2017;**3**(1): 89-94.
4. Alexopoulos, E.C.,. Introduction to multivariate regression analysis. *Hippokratia*, 2010;**14**(Suppl 1): p.23.
5. Cohen J, Cohen P, West SG & Aiken LS Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Third Edition. Lawrence Erlbaum Associates, Mahwah, NJ2003.
6. Diem Ngo, T.H., La Puente, C.A. (). The Steps to Follow in a Multiple Regression Analysis. SAS Global Forum .*Statistics and Data Analysis*.2012; Paper 333-2012: Pp 1-12.
7. Douglas, E.A. (2021). Introduction to linear regression analysis. John Wiley & Sons.
8. Draper, N.R., & Smith, H. (1992). Applied Regression Analysis, Second Edition. John Wiley & Sons, Inc. New York.
9. Draper, N.R., & Smith, H. (1998). Applied Regression Analysis, Third Edition. Wiley Interscience Publication, United States.
10. Hornik, K., Stinchcomber, M., White, H. Multilayer feedforward networks are universal approximators. *Neural Networks*.1989;**2**: 359-366.
11. Jean Gaudart, B. G.. Comparison of the performance of multi-layer perceptron and linear regression forepidemiological data. *Computational Statistics & Data Analysis*.2004;547-570.
12. M.W. Gardner, S.R. Dorling. Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. *Atmospheric Environment*,1998.**32**(14), 2627-2636.