IJRP.ORG
International Journal of Research Publications
ISSN: 2708-3578 (Online)

A Machine Learning Approach To Aid Diagnosis Of Patients High Blood Pressure Using Decision Tree Model

Omorogiuwa O[a], Idris A. A[b], Achuenu A[c]

[a] ask4osas@iuokada.edu.ng
[a,b]Department of Computer Science and Information Technology, Igbinedion University Okada, Nigeria
[c]Computer Science Department, Auchi Polytechnic, Edo State, Nigeria

Abstract

This study focus on utilizing the strength of Decision Tree Algorithm to develop a High Blood Pressure prediction model with the help of RapidMiner studio. A dataset containing 2000 records of patients with high blood pressure was collected from National Health and Nutrition Examination Survey (NHANES). The dataset was properly prepared and feature selection algorithms where used to determine the most relevant feature from the dataset. The features used in the course of the study are chronic kidney disease, adrenal and thyroid disorder, level of haemoglobin, genetic pedigree coefficient, age, alcohol conception, sex, BMI, and salt conception. The data set was split into two parts, training data set and testing data set. The training data set consist of 80% of the data while the testing data set contains 20% of the data. The decision tree model was trained with the training dataset and the developed model was applied to the test dataset. The performance of the model was further evaluated and produced an accuracy of 87% which shows that the Decision Tree algorithm can be effective on the prediction of High Blood Pressure.

Keywords: Machine Learning (ML); Decision Tree; RapidMiner Studio; Data Science; ML Algorithms

1. Introduction

On a worldwide scale, very few numbers of attempts have been made by researcher's to predict blood pressure abnormality using Machine Learning (ML) algorithms. Different researchers considered different features and dataset for high blood pressure prediction. Some of the most common machine learning algorithms used include K-Nearest neighbor, Naïve bayes, Logistic regression, artificial neural network and Fuzzy logic. Artificial intelligence, machine learning and deep learning has greatly be impactful in the field of medicine especially in diagnosis, treatment of analysis of many ailments, Amruta (2018).
Blood pressure is the force of blood pushing against the blood vessel walls. It is measured in millimeters of mercury (mm Hg). High blood pressure (HBP) also called hypertension occurs when the blood pressure in the

arteries is consistently higher than it should be. Hypertension is determined by two factors, the systolic pressure and diastolic pressure. The systolic pressure is the pressure in the arteries when the heart beats while the diastolic pressure is the pressure in the arteries when the heart rests between beats. The American College of Cardiology and the American Heart Association's guidelines published in 2017 defines high blood pressure as a reading greater than 130 mm Hg for systolic blood pressure or greater than 80 mm Hg for diastolic blood pressure. Dangers of high blood pressure includes stroke, heart failure, Heart Attack and Memory Loss. So far, patients with high blood pressure show no symptoms at all. Few people may experience headaches or shortness of breath, but these signs and symptoms are not specific. Research as shown that high blood pressure is caused by a lot of factors such as high salt consumption, smoking, alcohol consumption, lack of proper diet, stress, lack of exercise, Too much fat and sometimes age (Mayo Clinic, 2021). In order to reduce the tendency of high blood pressure, it is important to predict the likelihood of a patient having high blood pressure, early for proper medication and caution.

2. Research Objectives

This study aim to employ the modern use of artificial intelligence and machine learning to develop a decision tree model which can be used to predict the possibility of high blood pressure in a patient.
The specific objectives of the work are as follow:

i. Apply Feature Selection Algorithm to select attributes that are relevant to the prediction of high blood pressure.
ii. Formulate a Decision Tree model for the prediction of high blood pressure.
iii. Apply and evaluate the performance of the developed model in predicting high blood pressure relative to other machine learning algorithms.

3. Research Instrument

RapidMiner forms a core data science and machine learning application software that will be used in achieving our research objectives. RapidMiner is developed for the data scientist which provides an integrated development environment for data preparation, data analysis, machine learning, deep learning, text mining and predictive analytics. It is a tool that can be used to develop models in the area of artificial intelligence and big data analytics. It also support large varieties of machine learning algorithms such as logistic regression, linear regression, random forest, artificial neural network, k-nearest neighbors, time series, etc.

4. Review of Related Literature

Mayilvaganan et al, (2014) proposed a human blood pressure classification analysis model using fuzzy logic. Various factors such as age, gender, smoking, alcohol Consumption and body mass index (BMI) where considered for the purpose of detecting early increase in blood pressure. The model was implemented using fuzzy membership functions, however, the model was not validated with live datasets which limits the accuracy of the model. Fernando et al., (2020) worked on the use of Artificial Neural Network for predicting hypertension. The research focused on estimating the association among gender, race, BMI, age, smoking, kidney disease and diabetes in hypertensive patients. The dataset used for the study was obtained from the National Health and Nutrition Examination Survey from 2007 to 2020. The study shows that artificial neural network using back propagation can be effective in predicting High Blood Pressure with an accuracy of 73.2%. Enid et al., (2018) in a research also worked on a prediction model of blood pressure for telemedicine using Artificial Neural Network. The data used for this research where collected from the health and body conditions of 498 people. The features considered includes gender, age, BMI, smoking status, exercise level, alcohol consumption level, stress level, and salt intake level. At the end of the experiment, the model was evaluated using the training dataset and gave an accuracy of 90.25% and 88.52% for back propagation respectively. The model was also evaluated on a testing data set and gave an accuracy of 94.28 and 93.75% for back propagation radial basis function (RBF) respectively. However, the research faced some few limitations due to the size of the dataset used.

Monday et al (2020) proposed an ordinal logistics regression model to study the occurrence of high blood pressure. The study shows that increasing age, body mass index and high salt intake are the major risk factors of hypertension. Age however, was proven to have a very low significance to High Blood Pressure compared to BMI and Gender (Monday 2020). Amin et al., (2021) in their research paper conducted a comparative study between the use of Logistic Regression Model and Support Vector Machine for the prediction of High Blood Pressure. The dataset collected was used in the10th annual physionet/computer in cardiology challenge on predicting acute hypotension episode in the intensive care unit (ICU). The dataset consist of 95 patient’s record with high blood pressure which was divided into a training and test dataset. A total of 10 features were selected and used to train the linear regression model and Support Vector Machine Model. At the end of the research, the Liner Regression Model produced an accuracy of 80% while the Support Vector Machine produced an accuracy of 88% which shows that Support Vector Machine is more effective in the prediction of High Blood Pressure compared to Liner Regression. However, the reach also suffers from lack of large dataset used in training the model and as a result would lift the performance of the model if applied to a larger dataset.

Babajide et al., 2017, proposed the use of naïve bayes classifier for high blood pressure prediction. For the purpose of the study, data was collected from 52 patients undergoing treatment at a hospital located in the south-western part of Nigeria and was prepared with the help of the Waikato Environment for Knowledge Analysis (WEKA) software application. Juan (2017) designed an optimized fuzzy classifier for the diagnosis of blood pressure with a new computational method for expert rule optimization. Latifa (2020), applied the concept of machine learning for predicting hypertension. Satyanarayana et al., (2018) proposed the use of random forest algorithm in the prediction of high blood pressure based on anger, age and anxiety. However, anger and anxiety are temporary biological state of humans and as a result might not be accurate features for the prediction of high blood pressure. More static features such as gender, BMI and genetic coefficient have to be considered for a more accurate prediction. Some notable research findings was equally achieved by Muhammad (2020) where he used k-nearest neighbor and support vector machine for predicting ischemic heart disease.

In the medical environment during diagnosis, medical personnel would occasionally face complex problems involving decisions that lead to different outcomes. If the decision process involves many sequential decisions, then the decision problem becomes difficult to visualize and to implement. Decision trees are indispensable graphical tools in such settings. They allow for intuitive understanding of the problem and can aid in decision making. (Ishwaran, 2009). Decision tree also have some few disadvantages. The “divide and conquer” method is used by a decision tree and as a result, they perform well if a dataset has relevant attributes, but less when a dataset has many complex and irrelevant attributes (Maimon, 2012).

5. Methodology

To develop the proposed model historical cases of high blood pressure will be collected and prepared with the help of Rapidminer studio. The choice of using RapidMiner was due to it flexible and interactive user friendly interface. Other data science modeling software such as WEKA and MATHLAB has little or no support for data preparation and analysis. RapidMiner however provides data importation and support for data preparation. RapidMiner also provides support for data visualization, database and cloud integration, In order to develop the Decision Tree Model in Rapid Miner Studio, the following procedures was applied:

i. Collect the High Blood Pressure dataset for modeling and load it into RapidMiner Studio with the help of the import and export feature.
ii. Prepare the dataset, properly convert the data type of the attributes and remove missing values using the Rapid Miner Turbo Prep.
iii. Apply feature selection algorithm techniques to select relevant attributes for more accurate prediction with the help of the Feature Selection tools in Rapid Miner.
iv. Split the dataset into two, 80% for training the model and 20% for testing.
v. Apply the developed Decision Tree model to the test dataset using the Apply Model operator in Rapid Miner Studio.
vi. Measure the performance of the model using the Classification Performance Operator.

Data Collection and Description

The dataset used for this research was collected from the National Health and Nutrition Examination Survey (NHANES) results for the years 2007 - 2020. The dataset is available online at https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx.The dataset to be used consist of 2000 records of patients with 14 features as shown in the table below.

Table 1.0: Features and Data Type of the Dataset

S/N	Feature	Data Type	Min Value	Max Value
1	Sex	Integer	0	1
2	Smoking	Integer	0	1
3	Age	Integer	18	75
4	BMI (Body Mass Index)	Integer	10	50
5	Salt Content In Diet	Integer	22	49976
6	Level Of Stress	Integer	1	3
7	Alcohol Consumption Per Day	Integer	0	499

8	Adrenal And Thyroid Disorder	Integer	0	1
9	Genetic Pedigree Coefficient	Real	0	1
10	Level of Hemoglobin	Real	8.100	17.560
11	Physical Activity	Real	628	49980
12	Blood Pressure Abnormality	Integer	0	1
13	Pregnancy	Integer	0	1
14	Chronic Kidney Disease	Integer	0	1

The dataset collected consist of 14 features as shown in the Table 1.0. All these features are factors that contribute to high blood pressure. The dataset consist of a total of 2000 patients with 1008 Males and 992 Females within the age range of 18 - 75. Out of 1008 males, 470 males have an abnormal blood pressure while 538 males have a normal blood pressure. From a total of 992 females, 517 females have an abnormal blood pressure while 475 females have a normal blood pressure. The data also consist of 199 female who were pregnant at the time of the test with 103 having an abnormal blood pressure and 96 having a normal blood pressure. From the dataset collected, more insights where obtained on patients with Smoking habit and patients affected with chronic kidney disease, and adrenal and thyroid disorder as shown in the Table 2.0.

Table 2.0: Number of Patients affected by a disease, disorder and smoking

	Smoking Habits	Chronic Kidney Disease	Adrenal And Thyroid Disorder
Male Patients Affected	505	496	453
Male Patients Not Affected	494	512	555
Female Patients Affected	514	514	434
Female Patients Not Affected	487	478	558

- ## Data Preparation

Data preparation simply is the act of preparing data to reduce noise and error in the data and structuring the data in a much suitable manner for data science. Data cleaning helps in identifying and removing missing or unrelated data from a group of dataset. Data transformation helps to manipulate and transform data into a suitable and understandable format such as renaming columns, removing columns, changing column data type, joining two datasets and much more.

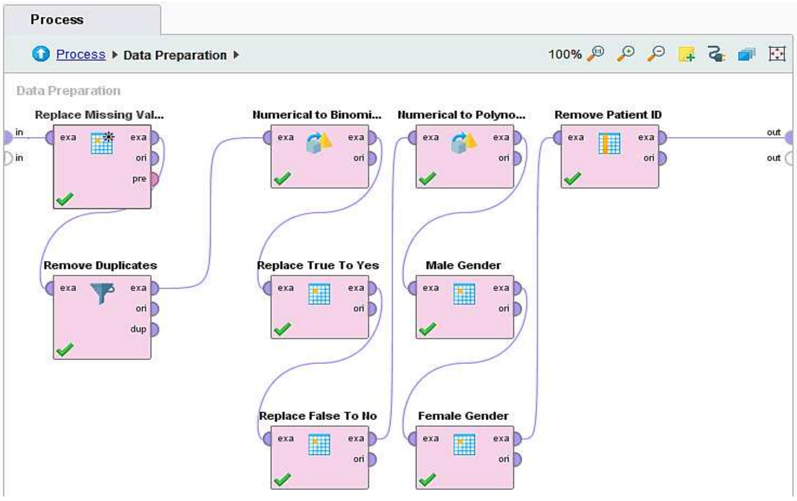

Figure 1.0: Data Preparation Process in Rapid Miner

The data preparation operations used to prepare the dataset collected shown in figure 1.0 is as follows

i. **Missing Values**: The data was first prepared to remove missing values from the data. The pregnancy attributes consists of two distinct values, 0 and 1 where 1 means true and 0 means false. The attribute consist of about 126 missing values which where all replaced by 0.

ii. **Duplicate Rows**: In order to improve the performance and learning curve of the prediction model, all duplicate rows on the data was filtered out.

iii. **Data Type Conversion:** The attributes adrenal and thyroid disorders, blood pressure abnormality, chronic kidney disease, pregnancy, and smoking all consist of integer values with only two distant values 1 and 0 which stands for true and false respectively. All the data contained in the attribute/column was converted in to a Binomial data where "True" replace "1" and "False" replaced "0". The data "True" and "False" was further replaced with "Yes" and "No" respectively for better understanding. The attribute "Gender" also consist of two distinct integer values 1 and 0. The column was further converted to a polynomial data type and the value "1" and "0" was replaced with "Male" and "Female" for better understanding of the data.

iv. **Prediction Label:** In order for the model to know which column to predict, the column has to be labeled. The column blood pressure abnormality for set at the prediction label for the data set.

v. **Remove Useless Attributes:** Part of preparing data is removing columns that are not relevant to the data. The dataset collected contains a column "Patient Id" which was used to number the records. The column was removed as it has no significant use for prediction.

Feature Selection

In order to reduce errors in prediction, it is necessary to select only attributes that are relevant. Feature selection algorithms are used in machine learning to select the most relevant attributes in a dataset. In the course of this study, we would be comparing the performance of three feature selection algorithms, weight by relief, weight by information gained and weight by gini index.

Model Development Procedure

To develop the proposed predictive model, the data is first collected, cleaned, normalized and prepared for developing the model. After the historical dataset has been cleaned and prepared, it would then be used to develop the proposed model. Figure 2.0 describes the process of developing the proposed model:

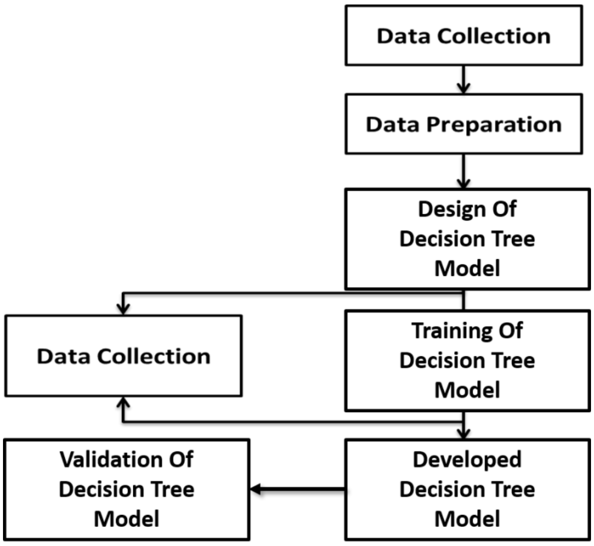

Figure 2.0: Model development process

The model development process is shown in figure 2.0. The prepared dataset is going to be divided into two parts, the first part for training the model and the second part for testing the performance of the model. The first part of the data is going to contain 80% of the overall data which would be used to formulate a Decision Tree Model. The second part of the data is going to contain 20% of the overall data which would be used to evaluate the performance of the model. RapidMiner provides an operator called **"Decision Tree"** which can be used to train and develop a Decision Tree Model from a set of training dataset. After the model has been developed, it can be applied to predict another given dataset with the help of the **"Apply Model"** operator in RapidMiner studio. Furthermore, the performance of the model would also be evaluated with the help of some operators in RapidMiner such as Performance operators, Split Validation and Cross-Validation.

6. Implementation, Result and Testing

In order to improve the performance of the model. Three feature selection algorithms Weight by Relief, Weight by Information Gained and Weight by Gini Index where used to select the most relevant features from the prepared dataset. The results obtained from the three feature algorithms as depicted in Table 1.0 shows that the features smoking, physical activity, pregnancy and level of stress are less relevant to the prediction of high blood pressure. The results also show that Chronic Kidney Disease, Adrenal and Thyroid Disorder, Level of Haemoglobin, Genetic Pedigree Coefficient, Age, Alcohol Conception, Sex, BMI and Salt

Conception are more relevant. It also shows that Chronic Kidney Disease, Adrenal and Thyroid Disorder, Level of Haemoglobin, and Genetic Pedigree Coefficient are the most important features in the prediction of high blood pressure. Table 1.0 shows the weight assigned to each attributes base on it relevance in the dataset within the range of 0 and 2.

Table 1.0: Result of the Feature Selection Algorithm Rated by Weight

S/N	Feature	Relief	Information Gained	Gini Index
1	Sex	0.013	0.002	0.002
2	Smoking	0.182	0.000	0.000
3	Age	0.049	0.005	0.003
4	BMI (Body Mass Index)	0.001	0.003	0.002
5	Salt Content In Diet	0.002	0.002	0.001
6	Level Of Stress	0.016	0.000	0.000
7	Alcohol Consumption Per Day	0.053	0.005	0.003
8	Adrenal And Thyroid Disorder	1.066	0.075	0.051
9	Genetic Pedigree Coefficient	0.155	0.066	0.041
10	Level of Hemoglobin	0.224	0.101	0.061
11	Physical Activity	-0.001	0.002	0.002
12	Pregnancy	-0.011	0.000	0.000
13	Chronic Kidney Disease	1.781	0.137	0.092

Based on the results obtained from the feature selection algorithm, only 9 out of the 13 features would be selected to build the prediction model, namely, chronic kidney disease, adrenal and thyroid disorder, level of haemoglobin, genetic pedigree coefficient, age, alcohol conception, sex, BMI, and salt conception.

- **Model Development**

The dataset used for the purpose of this study was collected from the National Health and Nutrition Examination Survey (NHANES) results for the years 2007 - 2020. The dataset consist of 2000 records of patients with 13 features and 1 prediction label. The dataset was further split into 2 parts, 80% for training the developed model and 20% for testing the performance of the developed model. The input data to the decision tree model include chronic kidney disease, adrenal and thyroid disorder, level of hemoglobin, genetic pedigree coefficient, age, alcohol Conception, sex, BMI, and salt conception (see Figures 4.0 and 5.0). The Decision Tree Model takes the inputs and further applies the Information Gained algorithms to classify the inputs and plot a graph of the tree based on the weight of the features as indicated in fig3.0.

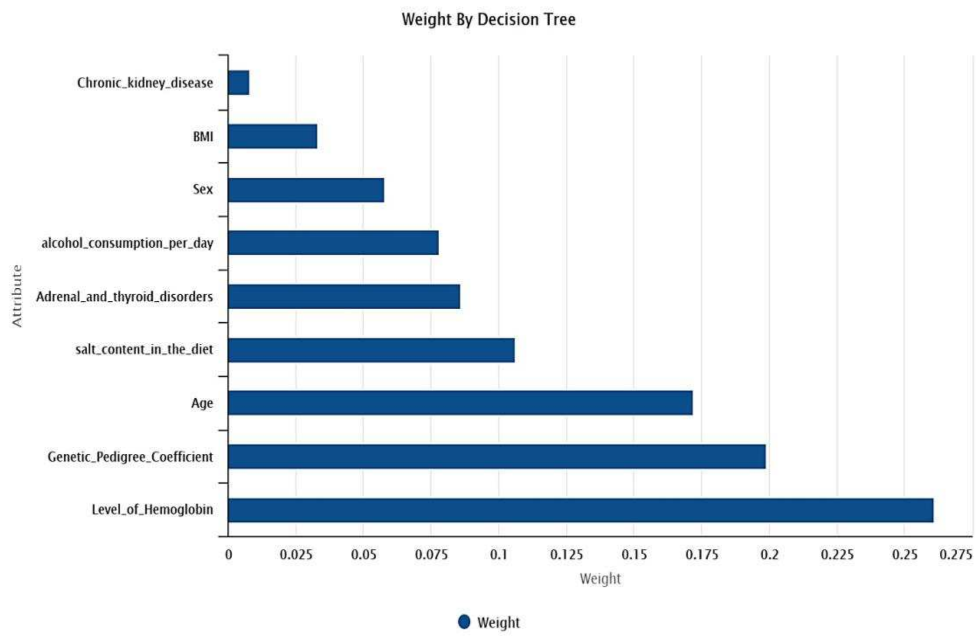

Figure 3.0: Weight Obtained By Decision Tree

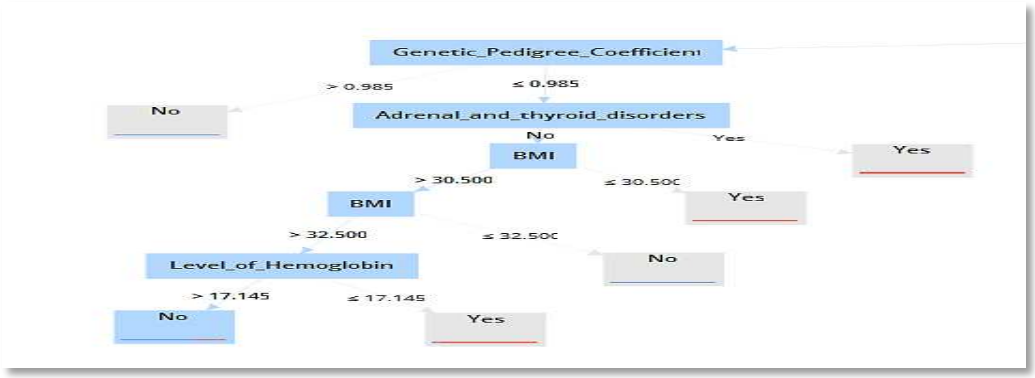

Figure 4.0: A section of the Decision tree model

Tree

```
Chronic_kidney_disease = No
|   Level_of_Hemoglobin > 14.550
|   |   Genetic_Pedigree_Coefficient > 0.985: No {No=2, Yes=0}
|   |   Genetic_Pedigree_Coefficient ≤ 0.985
|   |   |   Adrenal_and_thyroid_disorders = No
|   |   |   |   BMI > 30.500
|   |   |   |   |   BMI > 32.500
|   |   |   |   |   |   Level_of_Hemoglobin > 17.145: No {No=2, Yes=1}
|   |   |   |   |   |   Level_of_Hemoglobin ≤ 17.145: Yes {No=1, Yes=16}
|   |   |   |   |   BMI ≤ 32.500: No {No=2, Yes=0}
|   |   |   |   BMI ≤ 30.500: Yes {No=0, Yes=11}
|   |   |   Adrenal_and_thyroid_disorders = Yes: Yes {No=0, Yes=27}
```

Figure 5.0.: A section description of the Decision Tree Model

- **Model Application and Evaluation**

After training the Decision Tree Model with 80% of the prepared dataset, the model was applied to predict the 20% of the dataset which consist of 400 records. Out of 400 records, the model was able to classify 348 records correctly while 52 records where classified wrongly.

Table 2.0: Prediction result of the first 10 data

Row No.	Blood_Pressure_Abnormality	prediction(Blood_Pressure_Abnormality)	confidence(No)	confidence(Yes)
1	No	No	0.985	0.015
2	No	No	0.985	0.015
3	No	Yes	0	1
4	Yes	No	1	0
5	Yes	Yes	0	1
6	Yes	Yes	0.026	0.974
7	Yes	Yes	0	1
8	Yes	Yes	0	1
9	Yes	Yes	0	1
10	Yes	Yes	0.188	0.812

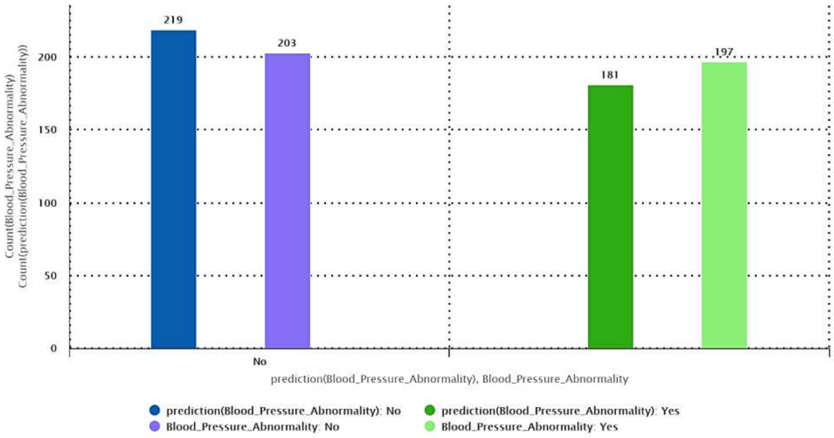

Figure 6.0: Bar Chart of the Model Prediction Result

In order to evaluate the performance of the prediction model, it is important to measure it accuracy. Accuracy of a classification model is measured using the below formula,

$$\boldsymbol{Accuracy} = \frac{\textbf{Total Number of correct prediction}}{\textbf{Total Number of prediction}} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}}$$

Where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

The True Positive (TP) is the number of positive outcome predicted correctly by the model. The False Positive (FP) is the number of positive outcome predicted wrongly by the model. The True Negative (TN) is the number of negative outcome predicted correctly by the model. The False Negative (FN) is the number of negative outcome predicted wrongly by the model.

Table 3.0: Decision Tree Model Prediction Outcome

	True No	**True Yes**
Prediction No	185	34
Prediction Yes	18	163

From the outcome of the prediction model, the true positive value is 163, the true negative value is 185, the false positive value is 34, and the false negative value is 18. By applying the formula, the accuracy of the prediction model in percentage was evaluated to be 87% as shown below.

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}} = \frac{\mathbf{163 + 185}}{\mathbf{163 + 185 + 34 + 18}} = \frac{\mathbf{348}}{\mathbf{400}} = \mathbf{0.87}$$

$$\boldsymbol{Accuracy\ in\ Persentage} = \mathbf{0.87} \times \mathbf{100} = \mathbf{87\%}$$

- **10-Fold Cross Validation**

Validation is one of the most important aspects of model development in data science. The cross-validation sometimes called rotation estimation is a method of validating the performance of a model. It is mainly used to estimate how accurately a model (learned by a particular learning Operator) will perform in practice. It divides a given dataset into 10 different set or folds, then hold out each fold in turn for testing and

IJRP.ORG
International Journal of Research Publications
ISSN: 2708-3578 (Online)

trains on the remaining 9 together which gives 10 evaluation results. The cross-validation test was carried out in RapidMiner studio and the average result is shown in the table below:

Table 4.0: 10-Fold Cross Validation Outcome

	True No	**True Yes**
Prediction No	923	227
Prediction Yes	90	760

From the outcome of the prediction model in Table 4.0, the true positive value is 760, the true negative value is 923, the false positive value is 227, and the false negative value is 90. By applying the formula, the accuracy of the prediction model in percentage was evaluated to be 84.15% as shown below.

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}} = \frac{\mathbf{760 + 923}}{\mathbf{760 + 923 + 227 + 90}} = \frac{\mathbf{1683}}{\mathbf{2000}} = \mathbf{0.8415}$$

$$\boldsymbol{Accuracy\ in\ Persentage} = \mathbf{0.8415} \times \mathbf{100} = \mathbf{84.15\%}$$

- **Model Performance Validation**

In the course of this study, a Decision Tree Algorithm was used to develop a High Blood Pressure prediction model with an accuracy of 84.15% from a 10 fold cross- validation. This section compares the result of the developed model with other machine learning algorithms. The machine learning algorithms to be compared with include Linear Regression, K-Nearest Neighbor, Naïve Bayes, and Random Forest.
Table 5.0 shows the results obtained from a 10- fold cross-validation using the entire prepared High Blood Pressure dataset.

Table 5.0: Model Performance Validation

S/N	**Machine Learning Algorithm**	**Accuracy**
1	Linear Regression	72.15%
2	K-Nearest Neighbor	48.90%
3	Naïve Bayes	82.40%
4	Random Forest	56.10%

The same dataset used for the development of the Decision Tree Model was used to develop another prediction model using Linear Regression, K-Nearest Neighbor, Naïve Bayes and Random Forest. The accuracy of the model was evaluated to be 72.15%, 48.90%, 82.40%, and 56.10% respectively for Linear Regression, K-Nearest Neighbor, Naïve Bayes and Random Forest. This proves that the Decision Tree algorithm is more effective on the prediction of High Blood Pressure with an accuracy of 84.15% compared to all other classification algorithms.

7. Conclusion

Accuracy has been the main issue of developing predictive model over the past few years. Various researchers have applied different methods to High Blood Pressure prediction like machine learning algorithms such as Support Vector Machine, Linear Regression, Random Forest, and Artificial Neural

Network. All this is done in attempt to efficiently predict the likelihood of high blood pressure and produced minimum accuracy. The research has discussed the application of Decision Tree in predicting High Blood Pressure. A dataset containing 2000 records of patients with high blood pressure was collected. The data set was spilt into two part, training data set and testing data set. The training data set consist of 80% of the data while the testing data set contains 20% of the data. The decision tree model was trained with the training dataset and the developed model was applied to the test dataset which produced an accuracy of 87%. The 10-fold cross validation was also used to evaluate the model and produced an accuracy of 84.15%. Furthermore, the performance of the model was compared to four other machine learning algorithm and was evaluated to be 72.15%, 48.90%, 82.40%, and 56.10% respectively for Linear Regression, K-Nearest Neighbor, Naïve Bayes and Random Forest which proves that the Decision Tree algorithm is more effective on the prediction of High Blood Pressure with a better accuracy of 84.15%.

References

Akhil M., Jabbar, B. D. (2013). Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm . International Conference on Computational Intelligence: Modeling Techniques and Applications , 85-94.

Amin Janghorbani, A. A. (2021). Prediction of Acute Hypotension Episodes Using Logistic Regression Model and Support Vector Machine: A Comparative Study. Retrieved from Research Gate: www.researchgate.net/publication/312117373

Amruta Powar, P. D. (2018). Heart Disease Prediction System Using Naïve Bayes Data Mining Technique. World Journal of Engineering Research and Technology (WJERT), 313-318.

Babajide O. Afeni, T. I. (2017). Hypertension Prediction System Using Naive Bayes Classifier. Journal of Advances in Mathematics and Computer Science, 1-11.

Daniel LaFreniere, F. Z. (2019, September 18). Using Machine Learning to Predict Hypertension from a Clinical Dataset. Retrieved from Research Gate: https://www.researchgate.net/publication/313804489_Using_machine_learning_to_predict_hypertension_from_a_clinical_dataset

Enid Wai-Yung Kwong, H. W.-H. (2018). A prediction model of blood pressure for telemedicine. Health Informatics Journal, 227-244.

Fernando López-Martínez, E. R.-V.-D. (2020). An artifcial neural network approach for predicting hypertension using NHANES data. Sanitas USA: Springer Nature.

Ishwaran, H. (2009). Introduction Decision Tree Classification. Retrieved from Miami: https://web.ccs.miami.edu/~hishwaran/papers/decisionTree_intro_IR2009_EMDM.pdf

Juan Carlos Guzman, P. M.-A. (2017). Design of an Optimized Fuzzy Classifier for the Diagnosis of Blood Pressure with a New Computational Method for Expert Rule Optimization. Algorithms, 1-27.

Latifa A. AlKaabi, L. S. R. (2020). Predicting hypertension using machine learning: Findings from Qatar Biobank Study. Retrieved from Plos One: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240370

Maimon, L. R. (2012). Decision Trees, Data Mining and Knowledge Discovery Handbook. Department of Industrial Engineering, Tel-Aviv University.

Mayilvaganan, M. a. (2014). Human Blood Pressure Classification Analysis Using Fuzzy Logic Control System in Data Mining. International Journal of Emerging Trends and Technology in Computer Science, 305-306.

Mayo Clinic. (2021). High Blood Pressure (Hypertention). Retrieved from Mayo Cliinic: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410

Monday Osagie Adenomon, a. D. (2020). Modelling Hypertension and Risk Factors among Adults Using Ordinal Logistics Regression Model. Retrieved from Preprints: https://www.preprints.org/manuscript/202001.0291/v1

Muhammad Yaqoob, F. I. (2020). Comparing predictive performance of k-nearest neighbors and support vector machine for predicting ischemic heart disease. Research Journal in Advanced Sciences, 49-60.

Norvig, S. J. (2009). Artificial Intelligence, A Modern Approach. Upper Saddle River, New Jersey: Pearson Education, Inc.

Satyanarayana Nimmala, Y. R. (2018). High blood pressure prediction based on AAA++ using

Machine-learning algorithms. Cogent Engineering 5(1), p 1-12 https://www.tandfonline.com/doi/full/10.1080/23311916.2018.1497114