European Journal of Computer Science and Information Technology
Vol.9, No.4, pp.42-53, 2021
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)

COMPARATIVE ANALYSIS OF SELECTED MACHINE LEARNING ALGORITHMS BASED ON GENERATED SMART HOME DATASET

Musa Martha Ozohu
Department of Computer Science
University of Port Harcourt
Port Harcourt, Nigeria.

Oghenekaro Linda Uchenna
Department of Computer Science
University of Port Harcourt
Port Harcourt, Nigeria.

ABSTRACT: *There has being recent interest in applying machine learning techniques in smart homes for the purpose of securing the home. This paper presents the comparative study on six classification algorithms based on generated smart home datasets. These includes Logistics Regression, Support vector machine, Random forest, K-Nearest Neighbor, Decision Tree and Gaussian Naïve Bayes. Two different smart home datasets were generated and used to train and test the algorithms. The confusion matrix was used to evaluate the outputs of the classifiers. From the confusion matrix, Prediction Accuracy, Precision, Recall and F1-Score of the models were calculated. The Support Vector Machine (SVM) outperformed the other algorithms in terms of accuracy on both datasets with values of 67.89 and 88.56 respectively. The SVM and Logistics Regression also maintained the highest precision of 100.0 as compared to the other algorithms*.

KEYWORDS: smart home, classification algorithms, support vector machine

INTRODUCTION

Algorithms are the bedrocks for problem solving. In data science, classification algorithms are supervised machine learning techniques used to categorize data into a class or category based on a training dataset. They are referred to as supervised because a training dataset is given. In classification, a program learns from the given dataset and then classifies new observations into a number of classes or groups.

Different types of classification algorithms exist and each of them have their advantages and disadvantages and are best suited for different purposes, hence, the need to analyze and compare them so as to know where to apply them to get optimal results. The most common classification algorithms in machine learning in no specific order are; Logistic Regression, Naïve Bayes, K-Nearest Neighbors, Decision Tree and Support Vector Machines (Mohssen *et al*., 2016). In this paper, a comparison of these six algorithms was done based on some performance metrics.

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

REVIEW OF RELATED WORKS

These literatures on the comparison and analysis of some machine learning algorithms were read and reviewed in this research:

Deepika *et al.,* (2018), applied four machine learning algorithms for the detection of dementia which is a neurodegenerative disorder. The machine learning algorithms applied are J48, Naive Bayes, Random Forest and Multi-layer perception. Of all these algorithms J48 outperformed other algorithms with an accuracy of 99.52% on Oasis cross sectional data and 99.20% on Oasis longitudinal data.

Pranckevičius and Marcinkevičius, (2017) investigated Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression classifiers implemented in Apache Spark, i.e. the in-memory intensive computing platform. In experiments, short texts for product-review data from Amazon1 were analyzed. Based on the size of training data sets, and the number of n-grams, all the classifiers mentioned above were compared based on their classification accuracy. Their findings indicate that Logistic Regression multi-class classification method for product-reviews has the highest (min 32.43%, max 58.50%) classification accuracy in comparison with Naïve Bayes, Random Forest, Decision Tree, and Support Vector Machines classification methods. On the contrary, Decision Tree has got the lowest average accuracy values (min in trigram: 24.10%, max in uni/bi/tri-gram: 34.58%). Also, their investigation indicates that increasing the size of the training data set from 5000 to 75000 reviews per class leads to insignificant growth of the classification accuracy (1 – 2%) of Naïve Bayes, Random Forest, and Support Vector Machines classifiers. These results show that a training set size of 5000 reviews per class is sufficient for all analyzed classification methods, and classification accuracy relates more to the n-gram properties.

Vandana *et al.,* (2021), investigated the utilization of the growth of regularly generated data from many financial activities by means of an automated process. The automated process was developed by using machine learning based techniques that analyze the data and gain experience from the underlying data. Different important domains of financial fields such as Credit card fraud detection, bankruptcy detection, loan default prediction, investment prediction, marketing and many other financial models were modeled by implementing machine learning models. Two parametric models namely Logistic Regression, Gaussian Naive Bayes models and two non-parametric methods such as Random Forest, Decision Tree were implemented in this paper. The performance of each classifier on each considered domain was evaluated by various performance metrics such as accuracy, recall, precision, F1-score and mean squared error. In the credit card fraud detection model the decision tree classifier performed the best with an accuracy of 99.1% and, in the loan default prediction and bankruptcy detection model, the random forest classifier gave the best accuracy of 97% and 96.84% respectively.

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

Venkata and Shaik, (2020), emphasized the implementation of different machine learning algorithms for network-based intrusions, analyses data imbalance and its impact on classification and anomaly detection. A pair of balanced and imbalanced datasets, NSL-KDD and CICIDS were considered as benchmark datasets for evaluation. Random Forest classifier is used to determine the best set of features for feature selection. The set of supervised and unsupervised algorithms used for the implementation included - K-Nearest Neighbors, Naive Bayes, Random Forest, Decision Trees, K-Means, Logistic Regression, Isolation Forest, and Local Outlier Factor. Implementation results indicated that in case of supervised learning, Random Forest outperformed the other methods, whereas K-Means performed better than other unsupervised learning methods.

Sylwia *et al.,* (2021), elaborated on how text analysis influences classification, which is a key part of the spam-filtering process. Three machine-learning methods allowing a user to classify e-mails as desirable (ham) or potentially harmful (spam) messages were compared in the paper to illustrate the operation of the meta-algorithm. Classifiers such as k-nearest neighbors (k-NNs), support vector machines (SVM), and the naïve Bayes classifier (NB) were used in this research. The conducted research gave the conclusion that multinomial naïve Bayes classifier can be an excellent weapon in the fight against the constantly increasing amount of spam messages. It was also confirmed that the proposed solution gives very accurate results.

Sayali and Channe, (2016), comparatively reviewed various classification techniques, hence their advantages and disadvantages. From survey and analysis on comparison among data mining classification algorithms (Decision tree, KNN, Bayesian), it showed that all Decision Tree's algorithms are more accurate and they have less error rate and they are easier algorithms as compared to K-NN and Bayesian. The knowledge in decision tree is represented in the form of [IF-THEN] rules which is easier for humans to understand. The result of implementation in WEKA on the same dataset showed that Decision Tree outperformed the other algorithms and Bayesian classification having the same accuracy as that of decision tree while K-NN does not give good results. However, the comparative study showed that each algorithm has its own set of advantages and disadvantages as well as its own area of implementation. None of the algorithm can satisfy all constrains and criteria.

Shler *et al.*, (2018), presented a performance comparison among these classifiers: Support Vector Machine (SVM), Logistics Regression (LR), K-Nearest Neighbors (K-NN), Weighted K-Nearest Neighbors (Weighted K-NN), and Gaussian Naïve Bayes (Gaussian NB) with an intent to improve the accuracy of breast cancer classification using data mining techniques. The dataset was taken from UCI Machine Learning Repository. The focus of this study was to classify breast cancer in women using the application of machine learning algorithms based on their accuracy. The results have revealed that Weighted K-NN (96.7%) has the highest accuracy among all the classifiers.

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

Kartik *et al.,* (2021), implemented machine learning algorithms such as Logistic Regression, Naïve Bayes, Random Forest, K-Nearest Neighbor, Gradient Boosting, Support Vector Machine, and Neural Network algorithms for the detection of fraudulent transactions. The objective of this paper was to detect fraudulent credit card transactions over non-fraudulent transactions and to use machine learning algorithms to predict fraud efficiently and accurately. A comparative analysis of these algorithms were performed to identify an optimal solution. The analysis showed that of the various Machine Learning algorithms implemented, the Gradient Boosting algorithm outperformed the other algorithms with an accuracy of 95.9%, followed by the Support Vector Mechanism with an accuracy of 94.7%.

Haitham *et al.,* (2020), aimed at using the Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) for univariate time series prediction. The goal of this study was to determine which algorithm performs better between Support Vector Machine and K-Nearest Neighbor in predicting time series data. The dataset for the monthly gold prices was used during the period from Nov-1989 – Dec-2019, which represents 362 observations. The results from the models were compared, it was observed that results from SVM were more accurate (with the lowest Root median standard deviation, RMSD). SVM and K-NN models were fitted based on 90% of data as training set, and then their accuracy was compared using the statistical measure RMSE.

Ratna and Sharavari, (2018), developed predictive models using eight machine learning algorithms namely Logistic Regression, K Nearest Neighbors (KNN), Support Vector Machines, Gradient Boost, Decision tree, MLP, Random Forest and Gaussian Naïve to predict the population who are most likely to develop diabetes on Pima Indian diabetes dataset. The studies revealed that Logistic Regression and Gradient Boost classifiers achieve higher test accuracy of 79 % compared with the other classifiers.

Isaac *et al.,* (2021), examined the performance of twenty-one (21) MLAs for classification and regression tasks based on six datasets from different domains. Empirically they compare their prediction results based on accuracy, balanced accuracy, F1-score, Area Under the Curve (AUC), root mean square error, r-squared and adjusted r-squared. The random forest algorithm gave a consistent performance across all the six datasets in classification and regression tasks. However, on average, XGBoost outperformed all Machine language algorithms (MLAs) applied in the research. The dummy algorithm and the linear regression were more moderate than the rest of the applied MLAs in computational complexity. Nevertheless, the overall study outcome shows that MLAs algorithms efficiently solve everyday challenges in elections outcome, financial fraud, network intrusion detection, meteorological forecast and heart diseases discovery; but their performance varies across domains and dataset dimensions.

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

MATERIALS AND METHODS

The datasets for the comparative analysis of these algorithms were generated from the rules spelled out in Table 1. The activities were generated from a smart home prototype which was from our previous research. The rules were generated based on the plan of the smart home. Certain activities can only precede others.

Table 1 has four columns; serial number, activity, activity meaning and next possible activity. In serial number 1, the activity HEN meaning Home Entry can only be followed by the activities in the 'next possible activity' column which are KEN, REN, HEX, SNG, DNG and CAY. For serial number 2, the activity KEN can only be followed by the activities CNG and KEX. The same follows for all the other rows of the table.

Table 1: Rules for constructing the user behaviour dataset

S/N	ACTIVITY	ACTIVITY MEANING	NEXT POSSIBLE ACTIVITY
1.	HEN	Home Entry	KEN, REN, HEX, SNG, DNG, CAY
2.	KEN	Kitchen Entry	CNG, KEX
3.	KEX	Kitchen Exit	SNG, CAY, DNG, REN, HEX, ENG
4.	CNG	Cooking	HEX
5.	HEX	Home Exit	HEN
6.	REN	Rest Room Entry	LNG, BNG, REX
7.	LNG	Stooling	BNG, REX
8.	BNG	Bathing	LNG, REX
9.	REX	Rest Room Exit	SNG, CAY, DNG, HEX
10.	SNG	Sleeping	KEN, REN, CAY, DNG, HEX
11.	CAY	Clothing Activity	DNG, HEX, SNG, REN, KEN
12.	DNG	Studying	HEX, CAY, REN, HEN, SNG
13.	ENG	Eating	KEN, SNG, REN, CAY, HEX

Using the rules above, two different sets of dataset were generated. The 5-feature dataset as shown on table 2 and the 7-feature dataset as shown on table 3. For the 5-feature dataset, the input features are Weekday (i.e day of the week), day hour (time of the day), current activity (the activity taking place), previous1activity (the activity that took place before the current one) and previous2activity (the activity that took place before previous1activity). The output or classification which is termed Normality in this research is either Yes or No which is represented on the last column of the table. On the first row, on a Sunday at about 12 midnight, current activity is DNG (studying), previous1activity is KEX (Kitchen Exit) and previous2activity is BNG (Bathing). These three sequence of activities form what we refer to as user behaviour which is expected to be predicted as No according to the training dataset. For the second row, on Sunday, at about 3am, the current activity is SNG (Sleeping), previous1activity is LNG (Stooling) and previous2activity is REN

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

(Rest Room Entry). These sequence of activities is expected to be predicted as Yes according to the training dataset. The same explanation follows for all the other rows on the table.

Table 2: 5-Feature Dataset

Weekday	Day Hour	Current Activity	Previous 1 Activity	Previous 2 Activity	Normality
Sunday	0	Dng	Kex	Bng	No
Sunday	3	Sng	Lng	Ren	Yes
Sunday	7	Hex	Cay	Kex	Yes
Sunday	9	Hen	Hex	Kex	Yes
Sunday	16	Kex	Ken	Eng	Yes
Monday	1	Dng	Hen	Eng	No
Monday	5	Lng	Ren	Kex	Yes
Monday	8	Eng	Kex	Cng	Yes
Monday	11	Hex	Rex	Cay	Yes

Table 3: 7-Feature Dataset

Weekday	Day Hour	Current Activity	Previous Activity 1	Previous Activity 2	Previous Activity 3	Previous Activity 4	Normality
Sunday	0	Dng	Kex	Bng	Ken	Sng	Yes
Sunday	3	Sng	Lng	Ren	Hex	Eng	No
Sunday	7	Hex	Cay	Kex	Cng	Ken	Yes
Sunday	9	Hen	Hex	Kex	Cng	Ken	Yes
Sunday	16	Kex	Ken	Eng	Ren	Hex	No
Monday	1	Dng	Hen	Eng	Cay	Sng	Yes
Monday	5	Lng	Ren	Kex	Cng	Ken	
Monday	8	Eng	Kex	Cng	Ken	Hen	Yes
Monday	11	Hex	Rex	Cay	Kex	Ken	No

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

The 7-Feature dataset has eight columns; the first seven columns are the input features into the model. These are Weekday, Day hour, Current Activity, Previous Activity 1, Previous Activity 2, Previous Activity 3 and Previous Activity 4 while the last column named normality is the outcome of the classification which is either a Yes or No. It is similar the 5-feature dataset, only that it has 2 additional columns (previousActivity3 and previousActivity4) which are not present in the 5-Feature dataset. For the first row on table 3, on Sunday morning at about 12 midnight, current activity is DNG (studying), previous1activity is KEX (Kitchen Exit), previous2activity is BNG (Bathing), previous3activity is KEN (Kitchen Entry) and previous4activity is SNG (sleeping). These series of activities is expected to be predicted as Yes according to the training dataset. On the second row, on a Sunday morning, at about 3am, currentActivity is SNG (Sleeping), previous1activity is LNG (Stooling), previous2activity is REN (Rest room Entry), previous3activity is HEX (home Exit) and previous3activity is ENG (Eating). These sequence of activities is expected to result in a No, according to the training dataset. No signifies anomaly detection. The same explanation follows for all the other rows of the table.

Three thousand rows of data were generated for each of the datasets and these two categories of datasets were used to train and test the models using the six different algorithms in order to evaluate the performance of the algorithms based on their accuracy, precision, recall and f1-score.

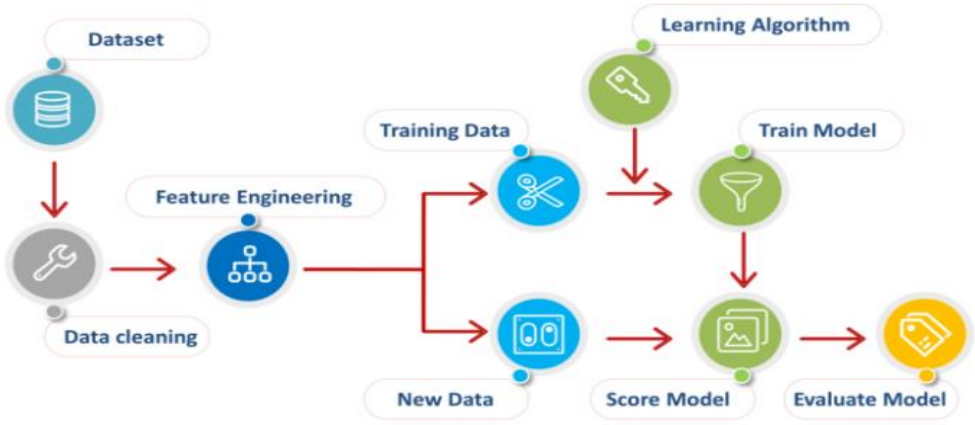

Figure 1: The Model building process

Data sets were saved in .csv format which stores tabular data in plain text. The dataset was imported, pandas library was imported to help manage the dataset and numpty library was also imported to perform mathematical functions and scientific computing. The machine learning algorithms to train the models were also imported. These includes Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest and Gaussian Naïve Bayes.

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

The dataset was cleaned and analyzed. Missing data were handled and categorical data were encoded. The dataset was then splitted into the testing and training sets. 70% of the dataset was used for model training while the remaining 30% was used for testing the models. The training dataset was fit into the various classifiers or algorithms to create or train the models. The testing dataset was used to run predictions. The predicted output was evaluated as against actual results and the confusion matrix was used as an evaluation metrics to calculate the performance of the models. From the result of the confusion matrix, prediction accuracy, precision, recall and f1_score for the models were further calculated. Accuracy is the sum of true positives and true negatives divided by the total number of samples. It is the proportion of the total number of predictions that were correct. Precision is the proportion of the predicted positives cases that were correct. It is the ability of the classification model to return only relevant instances. Recall is the ability of the classification model to identify all relevant instances. It is the proportion of positive cases that were correctly identified. F1_Score is the weighted average between precision and recall. Of all of these metrics, Accuracy of the model is of the highest priority to any programmer or developer.

RESULTS

The first set of models were developed using the 5-Feature dataset and the following results were obtained.

Table 4: Accuracy score for the models

	Score	Model
0	67.890000	Support Vector Machine
1	66.670000	Random Forest
2	66.440000	Logistic Regression
3	66.330000	K-Nearest Neighbor
4	65.000000	Decision Tree
5	48.222222	GaussianNB

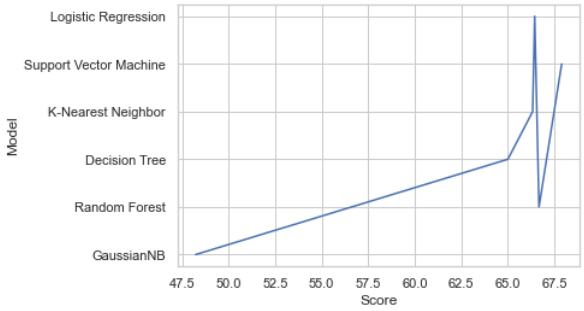

Figure 2: Accuracy score for the models

Table 5: Precision score for the models

	Score	Model
0	100.0	GaussianNB
1	83.0	Support Vector Machine
2	75.0	K-Nearest Neighbor
3	74.0	Logistic Regression
4	73.0	Random Forest
5	69.0	Decision Tree

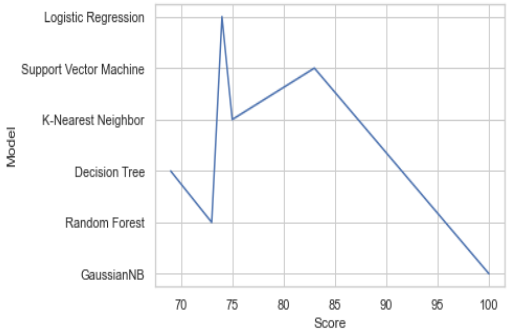

Figure 3: Precision score for the models

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

Table 6: Recall score for the models

	Score	Model
0	71.6	Decision Tree
1	67.0	Random Forest
2	64.2	Logistic Regression
3	63.2	K-Nearest Neighbor
4	55.7	Support Vector Machine
5	10.7	GaussianNB

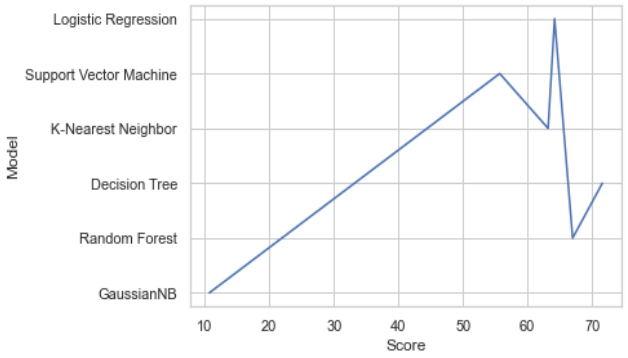

Figure 4: Recall score for the models

Table 7: F1 score for the models

	Score	Model
0	70.4	Decision Tree
1	70.0	Random Forest
2	68.9	Logistic Regression
3	68.5	K-Nearest Neighbor
4	66.8	Support Vector Machine
5	19.4	GaussianNB

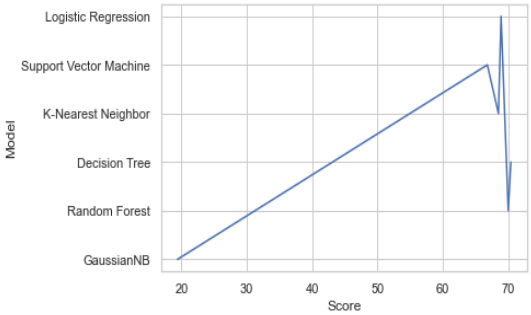

Figure 5: F1 score for the models

The second set of models were developed using the 7-feature dataset and the following results were obtained.

Table 8: Accuracy score for the models

	Score	Model
0	88.560000	Logistic Regression
1	88.560000	Support Vector Machine
2	87.330000	K-Nearest Neighbor
3	85.440000	Random Forest
4	81.890000	Decision Tree
5	76.333333	GaussianNB

Table 9: Precision score for the models

Figure 6: Accuracy score for the models

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

	Score	Model
0	100.0	Logistic Regression
1	100.0	Support Vector Machine
2	99.0	GaussianNB
3	96.0	K-Nearest Neighbor
4	87.0	Random Forest
5	77.0	Decision Tree

Table 10: Recall score for the models

	Score	Model
0	75.6	Decision Tree
1	73.6	Random Forest
2	70.4	Logistic Regression
3	70.4	Support Vector Machine
4	70.4	K-Nearest Neighbor
5	39.1	GaussianNB

Table 11: F1 score for the models

	Score	Model
0	82.6	Logistic Regression
1	82.6	Support Vector Machine
2	81.1	K-Nearest Neighbor
3	79.6	Random Forest
4	76.3	Decision Tree
5	56.1	GaussianNB

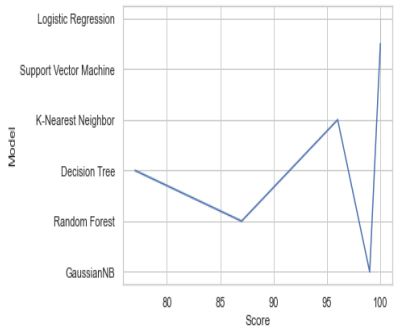

Figure 7: Precision score for the models

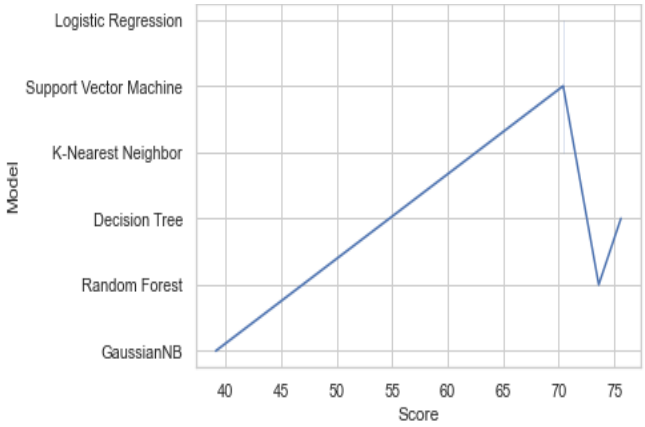

Figure 8: Recall score for the models

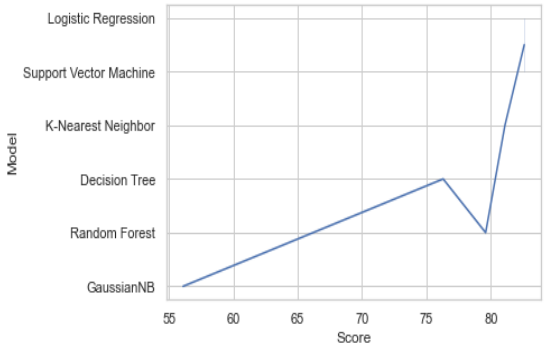

Figure 9: F1 score for the models

RESULT DISCUSSION

The results objectively show the performance of the selected machine learning algorithms when trained with the generated smart home datasets. As seen in table 4, Support Vector Machine (SVM), on the one hand outperforms all the other algorithms in terms of accuracy, with an

@ECRTD-UK- https://www.eajournals.org/
https://doi.org/10.37745/ejcsit.2013

accuracy score of 67.89, and this implies that SVM had most of its predictions correct, as when compared with the other algorithms. And on the other hand, Gaussian Naïve Bayes (GNB) recorded the least with, 48.22 and 76.33 accuracy score respectively. However, Gaussian Naïve Bayes recorded a precision score of 100 in table 5, and this implied that all behavioral patterns that were predicted to be positive were true positives. While the decision tree had the lowest precision value of 69.0. And in the reverse order, decision tree had the highest recall score of 71.6 which implies that of the total number of positive instances predicted by the decision tree algorithm was higher than positive instance prediction of other algorithms. Table 6 shows that the algorithm with F1 score closet to a perfect F-measure score was the decision tree with a value of 70.4, followed by the random forest till it got to the GNB with the least F1 score of 19.4. The 7-feature dataset was used to train all six machine learning algorithms, and same evaluation metrics were used to evaluate the algorithms. The Support Vector Machine (SVM) again, outperformed all the other algorithms in terms of accuracy with an accuracy score and 88.56, however the logistic regression model also had an accuracy score of 88.56, and then the Gaussian Naïve Bayes (GNB) still had the lowest accuracy score of 76.33. The SVM and logistic regression also maintained the highest precision score of 100.0 as compared to other algorithms. The decision tree still maintained the highest recall value of 75.6 and GNB had the lowest recall value of 39.1. The F1 score of logistic regression and support vector machine were same at a value of 82.6, and GNB recorded the lowest. Having carried out different simulations using two different datasets on all six machine learning algorithms, the evaluation results also showed that the dataset with higher number of features trained the algorithms more efficiently that the dataset with less features.

CONCLUSION

As accuracy is of highest priority among all performance metrics particularly to programmers and users, this paper therefore recommends the use of support vector machines as a classification algorithm to be used for binary classifications with smart home dataset.

It was also observed that the models built with the 7-feature dataset generally had higher performance scores when compared with the models built with the 5-feature dataset. This implies that the more the number of input features, the better the performance of the classification algorithms.

REFERENCES

Venkata, R. V. &Shaik, R. (2020). A Comparative Evaluation of supervised and unsupervised algorithms for Intrusion Detection. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4), 4834-4843. https://doi.org/10.30534/ijatcse/2020/93942020

Haitham, F., El, H. A. R &Amal, M. A. F. (2020). Comparison between support vector machines and K-nearest neighbor for time series forecasting. Journal of Mathematical and Computational Science, 10(2020), 2342-2359. DOI:10.28919/jmcs/4884.

Shler, F. K., Adnan, M. A. & Amira, B. S. (2021). A Comparative Analysis and Predicting for Breast Cancer Detection Based on Data Mining Models. *Asian Journal of Research in Computer Science*, 8(4): 45-59.

Ratna, P. &Sharavari, T. (2018). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Diabetes. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(5), 3966-3975. DOI: 10.11591/ijece.v8i5.pp3966-3975

Kartik, M., Manthan, N., Preity, P., Riya, R. &Supriya, P. (2021). Credit Card Fraud Detection System. *International Journal of Recent Technology and Engineering (IJRTE),* 10(2), 158-162.

Rapacz, S., Chołda, P. &Natkaniec, M. (2021). A Method for Fast Selection of Machine-Learning Classifiers for Spam Filtering. *Electronics*, 10(2083), 1-23. https:// doi.org/10.3390/electronics10172083.

Pranckevičius, T. &Marcinkevičius, V. (2017). Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification. *Baltic J. Modern Computing,* 5(2), 221-232. http://dx.doi.org/10.22364/bjmc.2017.5.2.05

Sayali, D. J. &Channe, H. P. (2016). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR),* 5(1), 1842-1845.

Peng, K., Tang, Z., Dong, L. & Sun, D.(2021). Machine Learning Based Identification of Microseismic Signals Using Characteristic Parameters. *Sensors*, 21(6967), https:// doi.org/10.3390/s21216967.

Deepika, B., Rita, C., Kavita, K. & Poonam, G. (2018). Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia. *Procedia Computer Science*, 132(2018) 1497–1502.

Saranya, V. &Porkodi, R. (2018). A Study and Analysis of Decision Tree Based Classification algorithms using R. *International Journal of Research in Advent Technology*, 6(7), 1681-1688. www.ijrat.org

Mohssen M. Z. E., Muhammed B. K. & Eihab B. M. B. (2017). Machine Learning: Algorithms and Applications. CRC Press.