Original Article

Recognition of Idleness among Patients based on Activity using Random Forest over Decision Tree Algorithm.

Kamasani v Bharath Kumar[1], S. Magesh Kumar[2]

[1]Research Scholar, Department of Computer science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India Pincode: 602105.
[2]Project Guide, Department of Computer science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.
Email: kamasanivbk18@saveetha.com

Abstract

Aim:Prediction of human activity data using Random forest algorithm against Decision tree algorithm for better accuracy using machine learning. **Materials and Methods:** Random forest algorithm (N=20)and Decision Tree algorithm (N=20) these two algorithms are calculated by using 2 Groups and I have taken 20 samples for both algorithm and accuracy in this work. **Results:** Based on the Results Accuracy obtained in terms of accuracy is identified by Random forest algorithm (90.6%) over Decision Tree algorithm algorithm(86.66%).Statistical significance difference between Random forest algorithm and Decision Tree algorithm was found to be 0.001 ($p<0.05$), using $\alpha=0.05$ and power=0.80.conclusion: From the analysis it was observed that the human activity data using Random forest algorithm is better than Decision Tree algorithm for better accuracy using machine learning.

Keywords: Decision Tree algorithm, Random forest algorithm, machine learning, Prediction, accuracy, human activity.

DOI: 10.47750/pnr.2022.13.S04.217

INTRODUCTION

In this paper we are going to check the human activity of different people of different ages and we will check the accuracy using different machine learning models . In this experiment we are taking the daily activity data set and the experiment is being carried forward(Kamali 2017).The experiments have been carried out with a group of volunteers within an age bracket of 19-48 years. Each person performed different activities Upstairs, Downstairs, Sitting, Lying, Standing, Walking wearing a smartphone(Kamali 2017; Hernangómez et al. 2022). In the technology world the day to day activities are also very important for a person to know what he is doing in his life and this human activity can be used in different ways like in fitness centers,personalized smart watches and also in the healthcare department centers. Machine learning method is widely used for preferable results. In machine learning there are different types of methods used for predicting human activities(Labrador and Lara Yejas 2013) . In this project we are going to use the Random forest algorithm with the Decision Tree algorithm.As a result of the advancements in the field of medicine, our quality of life has dramatically improved, as evidenced by the rise in life expectancy. The rising costs of healthcare, especially for the elderly, have prompted various health care organizations to implement cost-cutting measures(Fu 2015).

In reference to the Prediction of human activities using machine learning .In Last 5 years 2017-2021 the Google Scholar has published more than 209 papers and the IEEE published more than 240 papers about the Prediction of human activity using machine learning . The analysis of Random forest algorithm and Decision Tree algorithm in high performance efficiency has been made using an experimental approach. My study opinion is the efficient prediction of human activity data using proposed algorithms against Decision Tree algorithms for better accuracy using machine learning(Ortiz 2015).

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020). Throughout the world directly or indirectly many people are using these devices for monitoring the day to day activities. Taking advantage of machine learning to recognize human activities is a widely discussed topic today. There has been a noticeable rise

in interest for implementing a wide variety of applications including health monitoring, indoor movements, navigation, and location-based services(Labrador and Lara Yejas 2013; Cook and Krishnan 2015). In a previous paper they experimented and the accuracy is obtained for Naive Bayes classifiers. The accuracy percentages are 77% and 89% respectively and in the same way many other algorithms are also used to finalize the accuracy of the model(Labrador and Lara Yejas 2013) .

In the existing system the prediction of human activity using Random forest algorithm performance is good as compared to the Decision Tree algorithm by using machine learning.

Material and methods

The study is done in the Department of Computer science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences and The sample size of 40 is classified into Random forest algorithm and Decision Tree algorithm each of 20 sample size, the pretest power of 0.8 , data set is gathered from kaggle and the data set name is Human Activity Recognition with Smartphones. The data set consists of information about the human activity of the people sitting ,standing etc(Jain and Murugesan 2021) .

In group 1 the Random forest algorithm is used with machine learning ,20 different values are taken from the human activity data set .Random forest algorithm has recently dominated machine learning and in kaggle() 2021) .The performance of the Random forest algorithm is accurate and speed .This algorithm Random forest algorithm a Supervised Machine Learning Algorithm mostly used in the regression problem that are widely classified. Generally the decision tree builds a tree like different samples and takes the most important and majority votes for the classification and takes the average in the case of the regression model(Benware 2021).

Algorithm for Random forest

1. Import the necessary libraries.
2. Import the train and test datasets.
3. Split the dataset into train and validation
   split_data = StratifiedShuffleSplit(n_splits = 1, test_size = 0.3, random_state = 10)
4. Define the main attribute for prediction from the dataset (Activity).
5. Ir_clf -> RandomForestClassifier()
6. Ir_clf.fit -> rf.fit(X_train, Y_train)c
7. score -> accuracy_score(Y_test,Y_p)
8. Print the predicted accuracy

In group 2 the Decision Tree algorithm is used with machine learning, 20 different values are taken from the human activity data set. Decision Tree algorithm has simple supervised machine learning and it is used widely in Classification and Regression problems , it is also simple to implement and work and the data set we can get in kaggle (Ahad, Antar, and Ahmed 2020;) 2021).The performance of the Decision Tree algorithm splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. The algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking(Hartmann 2011).

Algorithm for Decision Tree algorithm

1. Import the necessary libraries.
2. Import the train and test datasets.
3. Split the dataset into train and validation
   split_data = StratifiedShuffleSplit(n_splits = 1, test_size = 0.3, random_state = 10)
4. Define the main attribute for prediction from the dataset (Activity).
5. rf -> Decision Tree algorithm(n_estimators = 10)
6. rf -> rf.fit(X_train, Y_train)
7. score -> accuracy_score(Y_test,Y_p)
8. Print the predicted accuracy

To perform the study a testing setup of hardware configuration components consists of DELL i5 processor, RAM 8GB, HDD 1TB and software components consisting of Windows 10 OS, Google Collab, Chrome Browser, MS Excel. The process of testing, downloading the dataset from kaggle and feature extraction and cleaning process were made. To train the model, 70% of the dataset was considered and for test and validation 30% of data was

used. From the algorithm calculate the value of activity rows and columns .By using confusion matrix accuracy is calculated(Garfinkel and Grunspan 2018).

The dataset was collectioned from kaggle, in the data set it has records of 561 total number of attributes have 6 and categories Upstairs, Downstairs, Sitting, Lying, Standing, Walking(Goudie 2013).

Statistical Analysis

The statistical software which is used for doing analysis is IBM SPSS version 22(64bit) which is a analysis software which is done by uploading dataset to the software which gives the output as independent variables N,means,std.deviation,std.error mean with the precision as the output for given models Random forest algorithm, Decision Tree algorithm. IBM SPSS is used to compare Prediction of human activity data using Random forest algorithm against Decision Tree algorithm for better accuracy using machine learning.(Pallant 2020).

Results

Table 1 represents the Random forest algorithm(90.6%) method and grouped statistics were compared using group statistics for recorded data from simulation for 20 iterations (86.66%) for Decision Tree algorithm. In comparison, the Random forest algorithm has a high level of accuracy. This shows that the proposed algorithm has more accuracy compared to other algorithms. The standard deviation of Random forest algorithm is 1.142 and Decision Tree algorithm is .795 T-test for comparison for Random forest algorithm standard error mean of Random forest algorithm is 0.255 and for Decision Tree algorithm is 0.178. Table 2 shows the independent sample T-test comparison analysis, using sig (2 tailed) .001, mean difference of 4.400, std error difference is 0.311 and statistical significance is 0.016. Fig.1.Architecture for predicition of human activity recognition using Random forest over Decision tree. Fig.2. shows the bar graph created by plotting mean accuracy on the Y-axis and groups on the X axis. Graph clearly shows that the Random forest algorithm is more accurate than the Decision Tree algorithm. The error rate of the Decision Tree algorithm is lower than the Random forest algorithm.

Discussion

Based on the results obtained by independent T-test analysis, the significance value is determined. The significance value is 0.001 and the accuracy of the Random Forest algorithm is 90.06% is higher than the accuracy of the Decision Tree algorithm is 86.66%. The statistical significance in the study is 0.16 ($p<0.05$ Independent Sample T-Test) states that results in the study are significant(Bratko 1989).

In this project we are using machine learning for predicting the accuracy of the human activity using two different algorithms, both the algorithm helps in finding the accuracy of the human activity,Machine Learning is used for prediction because it can automatically improve through a data set and basically it's a computer algorithm which is very much useful for healthcare and also for the person to know how much time he spent in the day and can analyze his workouts or the set his day to day activities as per plan.A powerful technology has emerged to help elderly and differently-abled people(Benware 2021). Human Activity Recognition. Simple human activities have been identified and investigated so far. Recognizing complex human activities is still challenging(Nguyen et al. 2021).

Conclusion

The Random Forest algorithm is 90.06% and Decision Tree algorithm is 86.66%. Random Forest algorithm provides significantly more accuracy than the Decision Tree algorithm in predicting human activity using machine learning.

DECLARATION

Conflict of Interests

No conflict of interest in this manuscript.

Authors Contributions

Author KVBK was involved in data collection, data analysis and manuscript writing. Author SMK was involved in the conceptualization, data validation and critical review of manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

Personally thank the following organizations for providing financial support that enabled us to complete the study.

1. Soft Square Pvt.Ltd, Chennai
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

Reference

1. Ahad, Md Atiqur Rahman, Anindya Das Antar, and Masud Ahmed. 2020. *IoT Sensor-Based Activity Recognition: Human Activity Recognition*. Springer Nature.
2. Benware, Dewey. 2021. *Machine Learning Algorithms: How The Machine Learning Algorithms Work Behind The Scenes: Random Forest Algorithm Geeksforgeeks*.
3. Bratko, I. 1989. "Machine Learning." *Human and Machine Problem Solving*. https://doi.org/10.1007/978-1-4684-8015-3_10.
4. Cook, Diane J., and Narayanan C. Krishnan. 2015. *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*. John Wiley & Sons.
5. Fu, Yun. 2015. *Human Activity Recognition and Prediction*. Springer.
6. Garfinkel, Simson L., and Rachel Grunspan. 2018. *Computer Book*. Sterling Milestones.
7. Goudie, Andrew. 2013. *The Human Impact on the Natural Environment: Past, Present, and Future*. John Wiley & Sons.
8. Hartmann, Bastian. 2011. *Human Worker Activity Recognition in Industrial Environments*. KIT Scientific Publishing.
9. Hernangómez, Rodrigo, Tristan Visentin, Lorenzo Servadei, Hamid Khodabakhshandeh, and Sławomir Stańczak. 2022. "Improving Radar Human Activity Classification Using Synthetic Data with Image Transformation." *Sensors* 22 (4). https://doi.org/10.3390/s22041519.
10. Jain, Sarika, and San Murugesan. 2021. *Smart Connected World: Technologies and Applications Shaping the Future*. Springer.
11. Kamali, Mira. 2017. *Human Activity Recognition Using Smartphone Accelerometer and Machine Learning*.
12.) Konrad Massaron Banachewicz. 2021. *DATA ANALYSIS AND MACHINE LEARNING WITH KAGGLE: How to Win Competitions on Kaggle and Build A... Successful Career in Data Science*.
13. Labrador, Miguel A., and Oscar D. Lara Yejas. 2013. *Human Activity Recognition: Using Wearable Sensors and Smartphones*. CRC Press.
14. Nguyen, Binh, Yves Coelho, Teodiano Bastos, and Sridhar Krishnan. 2021. "Trends in Human Activity Recognition with Focus on Machine Learning and Power Requirements." *Machine Learning with Applications*. https://doi.org/10.1016/j.mlwa.2021.100072.
15. Ortiz, Jorge Luis Reyes. 2015. *Smartphone-Based Human Activity Recognition*. Springer.
16. Pallant, Julie. 2020. "Getting to Know IBM SPSS." *SPSS Survival Manual*. https://doi.org/10.4324/9781003117445-4.

TABLES AND FIGURES

Table 1. The Random Forest algorithm (90.06%) method and grouped statistics were compared using group statistics for recorded data from simulation for 20 iterations (86.66%)Decision Tree algorithm In comparison, the Random Forest algorithm has a high level of accuracy

	Algorithm	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	Random Forest	20	90.06	1.142	.255
	Decision tree	20	86.66	.795	.178

Table 2. Independent sample T-test was performed between Random Forest algorithm and Decision Tree algorithm to identify the significance and standard error determination between the algorithms are shown.

	Levene's Test for Equality of Variance	T-test for Equality of Means

		f	Sig	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence of the Differences	
									Lower	Upper
Accuracy	Equal variances assumed	6.33	0.16	14.13	38	.001	4.400	.311	3.770	5.030
	Equal variances not assumed			14.13	33.899	.001	4.400	.311	3.778	5.032

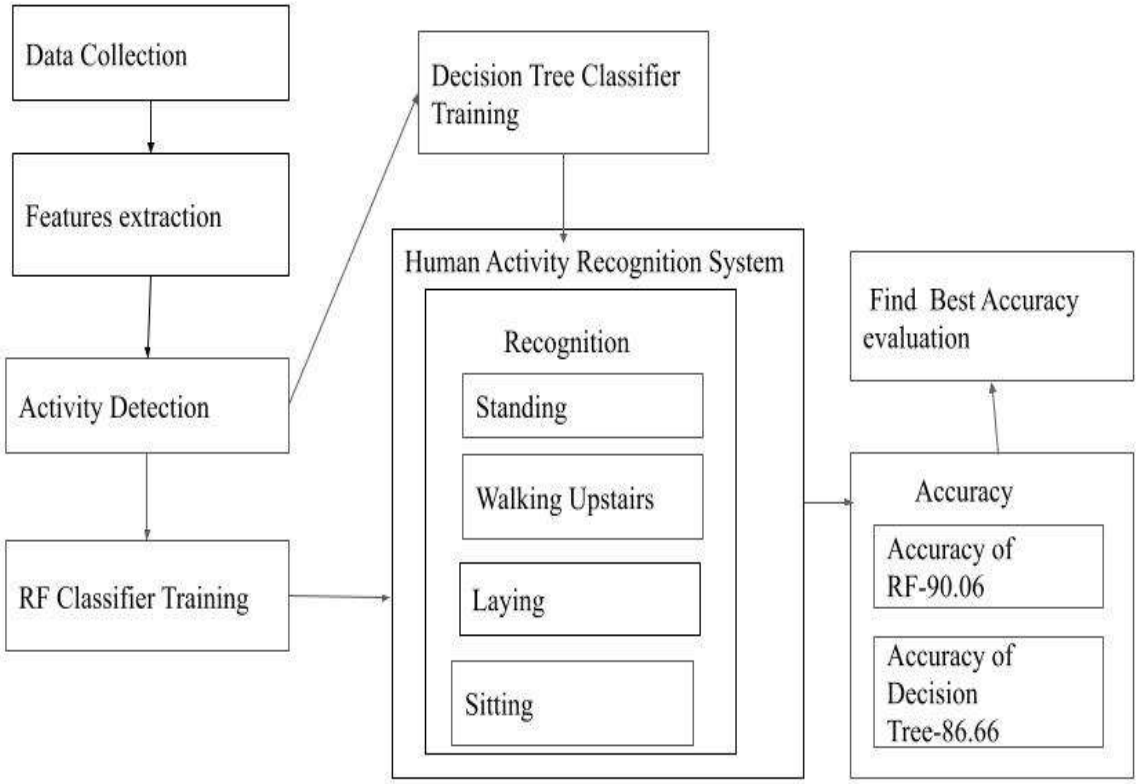

Fig.1.Architecture for predicition of human activity recognition using Random forest over Decision tree.

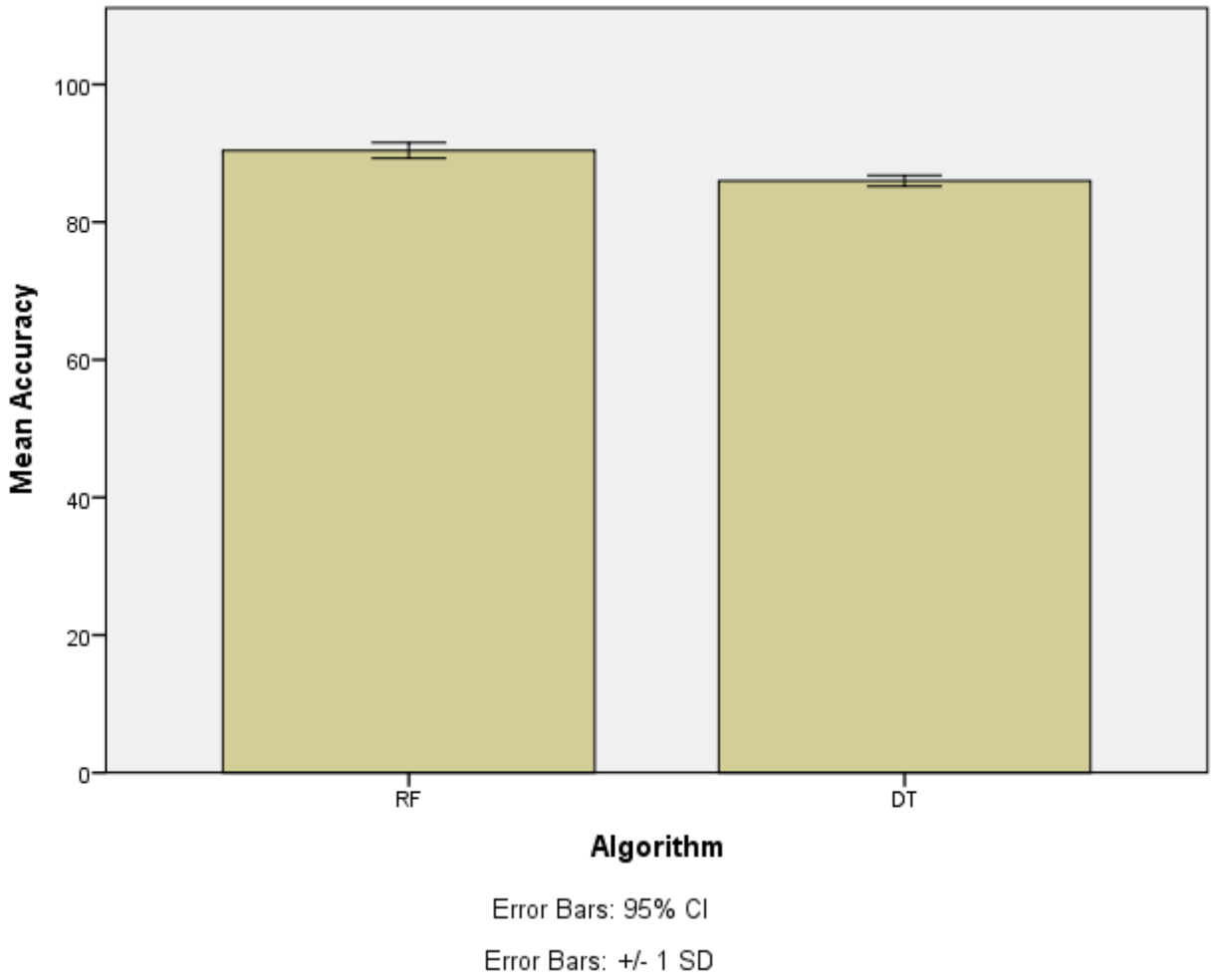

Fig. 2. Comparison of Random Forest algorithm and Decision tree algorithm in terms of mean and accuracy. The mean accuracy of the Random Forest algorithm is better than the Decision tree algorithmX-axis:Random Forest algorithm vs Decision tree algorithm, Y axis: Mean accuracy of detection ±1 SD.