IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A COMPARISON OF FEATURE SELECTION ALGORITHMS FOR HUMAN ACTIVITY RECOGNITION

[1]Thangapriya S, [2]Dr Nancy Jasmine Goldena

[1]Research Scholar (Reg No: 20211242282010), [2]Associate Professor

[1,2]Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), affiliated to Manonmaniam Sundaranar University, Tirunelveli.

ABSTRACT

Recognition of human action is one of the active research areas in Machine Learning (ML) for numerous applications. In Human Activity Recognition, feature analysis is a critical area of research for improving the classifier. For this investigation, the UCI dataset is used. This field contains lots of Feature Selection (FS) algorithms. FS enhances model performance while reducing model complexity and dataset dimensionality. As a result, the best features are identified using FS algorithms that are carefully chosen. The three FS algorithms that were used for comparison were Recursive Feature Elimination (RFE), Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF). Three specific classifiers are used to compare these FS techniques. Performance evaluation metrics properly test the algorithm's performance, revealing which is the greatest among the others. The investigation's outcomes are presented using the tools PYTHON and RSTUDIO.

Keywords: *Classification, Feature Selection, HAR, KNN, LASSO, Machine Learning, Multi-Layer Perceptron, Naive Bayes, Performance Evaluation, Random Forest, Recursive Feature Elimination, Support Vector Machine.*

1. INTRODUCTION

In today's modern world, the health industry has grown to be a massive development. Due to the massive rise in population, traditional health care is impossible to provide for everyone's needs. Instead of using traditional tools to monitor elders' wellness, a new method should be used. In the past, one or two staff would constantly supervise people in the hospital for their safety. It will be extremely difficult for everyone to keep an eye on the elderly. At this time, health care plays a significant role in a variety of circumstances. The smart activity recognition approach is being used to monitor the elderly peoples, which is the newest method of health care. HAR study allows us to track the movements of the elderly in order to see what they are doing and to keep track of elderly person's activities.

2. FEATURE SELECTION METHODS

Feature Selection (FS) is a method of selecting key features that improves performance of the model. There could be a lot of features in the dataset that aren't valuable or significant enough to incorporate in the model construction. Feature selection is not just a major challenge in pattern recognition, but it's also a major issue in the sensor-based activity recognition sector. The selection of appropriate features has a significant impact on recognition accuracy. Feature selection can help remove irrelevant or redundant information from a model, boosting accuracy and minimizing computing complexity [1]. Three FS algorithms were used in this comparative analysis shown in Figure 1.

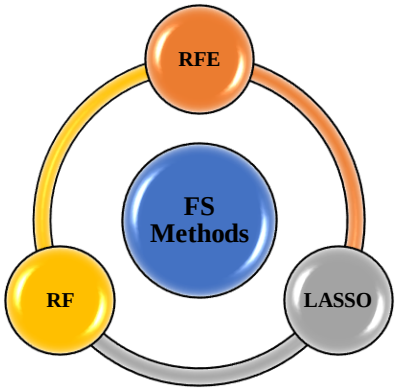


Figure 1: Feature Selection methods

3.1 Recursive Feature Elimination (RFE)

RFE is a FS technique that functions as a wrapper. This means that in the core of the approach, a machine learning algorithm is presented and used, which is wrapped by RFE and used to help selecting features. The ML model or a statistical method are used to score the features. This procedure is done until only a certain number of features are available [2]- [5]. To determine the best set of features RFE applies cross-validation to score multiple feature subsets and pick the finest set of features. The RFE process is basically a backward selection of prediction. This process starts by creating a model based on the whole list of predictors and assigning an importance value to each. The model is then reconstructed, and significance scores are calculated again after the least important predictor(s) are eliminated.RFE can be used to classify and predict outcomes. When using RFE, there are two vital configuration options: The number of features to pick and algorithm used to help choose features.

3.1.1 RFE Algorithm

STEPS	*DESCRIPTION*
Step 1	RFE uses a supervised learning estimator that has already been fitted using all aspects of the data.
Step 2	The coefficient associated with each characteristic is then considered, which is retrieved from the coef_ or feature importance_ attribute.
Step 3	The importance of these coefficients with respect to the target variable is represented by their values. The feature with the smallest absolute coefficient value is the least valuable.
Step 4	The least important coefficient is then removed from the list of features, and the model is rebuilt using the remaining features.
Step 5	It rebuilds the model with each iteration and removes the least significant feature(s), repeating the process until it reaches two features.
Step 6	The feature is then ranked depending on the time it took to eliminate it. The feature that was eliminated first has the highest rank, and so on.

3.2 Using LASSO for Valuable Selection

The Least Absolute Shrinkage and Selection Operator or LASSO, is a statistical algorithm that is used to select features and normalize data models. Selecting suitable features from the dataset available is one of the

most crucial steps. The built-in feature selection functionality of Lasso regression can be employed in a variety of situations. Exploratory data analysis would help in better understanding the most essential correlations between features and the target, enabling to choose the optimum model [6] -[8]. For the LASSO approach, the following equation is used.

$$\frac{1}{2N_{training}} \sum_{i=1}^{N_{training}} \left(y_{real}^{(i)} - y_{pred}^{(i)}\right)^2 + \alpha \sum_{j=1}^{n} |a_j|$$

Linear Regression with L1 regularisation is what Lasso Regression is. **Regularization:** Data points that act erratically or vary completely from the data's original distribution. **Outliers:** Can occur for a variety of reasons, including human mistake when creating the information, experimental error or fluctuation during exploratory survey.

3.2.1 LASSO Algorithm

STEPS	*DESCRIPTION*
Step 1	Prepare the dataset and use a cross-validation strategy to find the hyperparameter value.
Step 2	Cost function is used to train the model.
Step 3	When a feature is removed, its coefficient becomes zero.
Step 4	Recognize the main features.
Step 5	Classifiers are used to train the features.
Step 6	Display the results

3.3 Random Forest

One of the most common machine learning algorithms is random forests. They're so popular because they have strong predictive accuracy, low generalization error and are simple to follow in general. Random Forest feature selection falls under the area of Embedded methods. Filter and wrapper methods are combined together to form an embedded method. Algorithms with built-in feature selection methods are used to implement them.

2.3.1 RF Algorithm

STEPS	*DESCRIPTION*
Step 1	Prepare the dataset.
Step 2	Create a random forest classifier and train it.
Step 3	Make a new dataset with only those features called 'limited featured'.
Step 4	Train the classifier.
Step 5	Compare the 'full featured' classifier's accuracy against the 'limited featured' classifier's accuracy.
Step 6	Display the result

3. CLASSIFICATION METHODS

In this work three classifiers are used for classification. They are shown in **Figure2.**

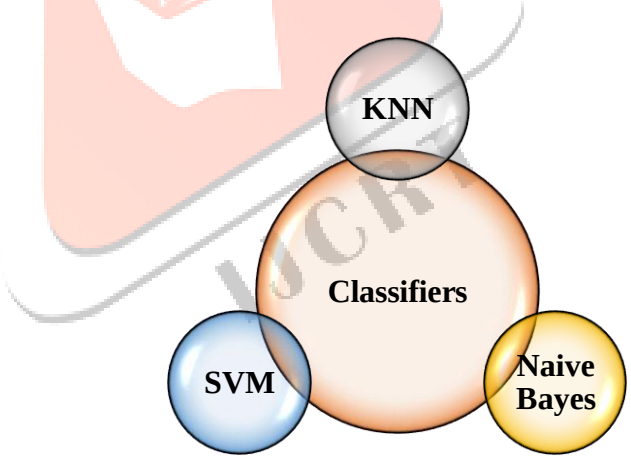


Figure2 Classification methods

4.1 KNN CLASSIFIER

The KNN method is one of the most fundamental Machine Learning algorithms. The KNN method stores all available data and classifies a new data point based on its similarity to the existing data. Which means that applying the KNN approach, fresh data can be efficiently classified into a well-defined category[11]- [12].

4.1.1 K-NN Algorithm:

STEPS	*DESCRIPTION*
Step 1	Choose K value.
Step 2	Calculate the Euclidean distance between each of the K neighbors.
Step 3	Using the estimated Euclidean distance, find the K closest neighbors.
Step 4	Count how many data points each category contains among these k neighbors.
Step 5	Assign the new data points to the category with the highest number of neighbors.
Step 6	The model is ready to go.

4.2 Naive Bayes

The Bayes theorem, often known as Bayes' Rule or Bayes' law, is a mathematical formula for calculating the probability of a hypothesis given previous information. It is conditional probability that decides outcome. It's a classification method that uses Bayes' Theorem and assumes predictor independence. A Naive Bayes classifier, in simple terms, proposes that the availability of one feature in a class is independent of the presence of any other feature. Naive Bayes is recognized to outperform even the most advanced classification systems due to its simplicity. The following equation will be used is naïve bayes classification [14].

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$

Where,

- P(A|B) is Posterior probability
- P(B|A) is Likelihood probability
- P(A) is Prior Probability
- P(B) is Marginal Probability

4.2.1 NB Algorithm

STEPS	*DESCRIPTION*
Step 1	Prepare the data set
Step 2	Using Naive Bayes to Improve the Training Set
Step 3	Predicting the outcome of a test
Step 4	Check the result's accuracy (Creation of Confusion matrix)
Step 5	Visualizing the results of the test set.

4.3 Support Vector Machine

SVM is a type of supervised machine learning approach that can be used for classification as well as regression. The goal of the SVM algorithm is to find a hyperplane in an N-dimensional space that classify data points clearly. The hyperplane's size is determined by the number of features [15].

4.3.1 SVM Algorithm:

STEPS	*DESCRIPTION*
Step 1	Choose the appropriate hyperplane.
Step 2	SVM Adjustment to the Training Set
Step 3	Predicting the outcome of a test
Step 4	Check the result's accuracy.
Step 5	Visualizing the results of the test set.

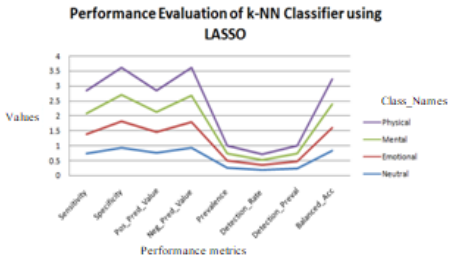
4. RESULT AND DISCUSSION

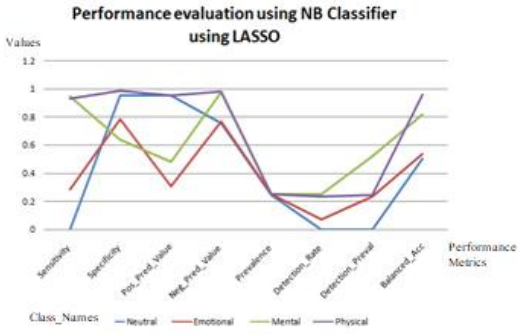
One of the most investigated areas in the world is classification problems. Classification Metrics assess a model's performance and tell you whether the classification is right or wrong, but each one does so in a unique way [19]. Using tree-derived feature significance to pick features for machine learning is a fairly simple, quick and generally accurate method of doing so [20],[21]. Because of their high accuracy, stability and ease of use, random forests are one of the most used machine learning algorithms. KNN is easy to use, sensitive to noisy training data and more successful when

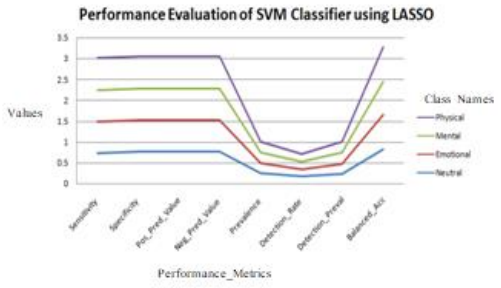
the training data is huge. However, KNN must continuously find the value of K, which might be difficult at times. The cost of computation is considerable since the distance between data points for all training samples must be calculated.

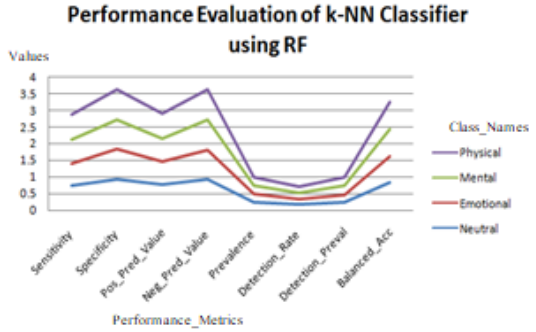
NB is a simple and quick method for predicting the test data set's class. In multiclass prediction, NB also performs well. When the assumption of independence is achieved, a Naive Bayes classifier outperforms alternative models such as logistic regression and requires less training data. In comparison to numerical input variables, NB performs well with categorical input variables (s). Naive Bayes is a quick and convenient machine learning algorithm for predicting a class of datasets. NB especially suitable for both binary and also multi-class classifications. In comparison to the other Algorithms, NB performs well in multi-class predictions. For text categorization tasks, NB is the most preferred option.Furthermore, when a categorical variable in the testing dataset has a class that was not included in the training set, the system will assign a zero probability and will be unable to predict. This is commonly referred to as "Zero Frequency." We can apply the smoothing approach to tackle this. Laplace estimation is one of the most basic smoothing techniques.

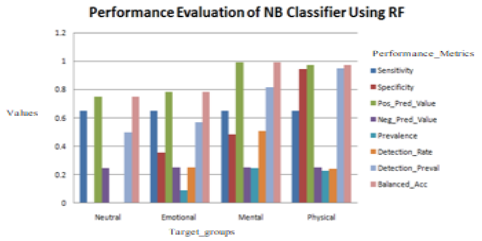
A linear hyperplane between these two classes is simple to create using SVM. However, another significant matter is whether we should need to manually implement this feature in order to have a hyper-plane. The kernel trick is a strategy used by the SVM algorithm. The SVM kernel is a function that converts a non-separable problem into a separable problem by taking a low-dimensional input space and transforming it to a higher-dimensional space. Simply told, it does some incredibly sophisticated data transformations before determining the best method for separating data based on labels. SVM is particularly successful in high-dimensional spaces with a clear separation boundary. When the number of dimensions is more than the number of samples, SVM is effective. SVM is memory efficient since it only utilizes a fraction of training points in the decision function. However, SVM does not perform well when dealing with large data sets because the training time is longer. When the data set has more noise, i.e. target classes overlap, SVM does not perform well. Probability estimates are produced using an expensive five-fold cross-validation method, which is not directly provided by SVM.

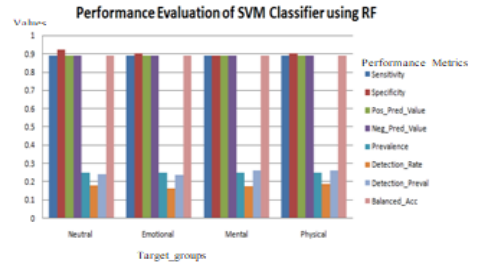


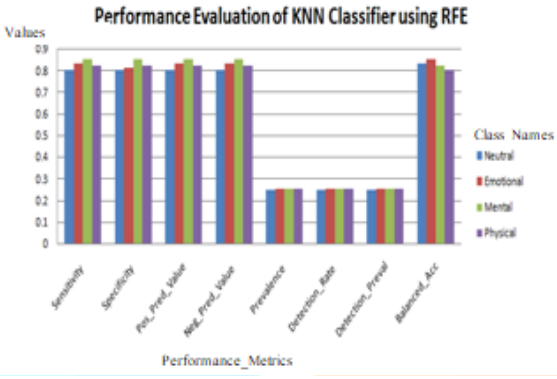


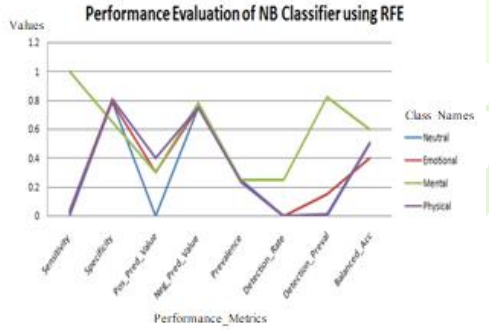


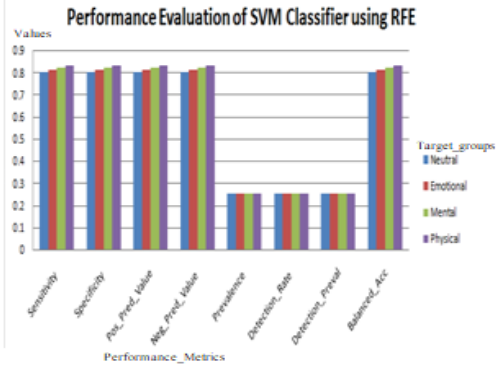


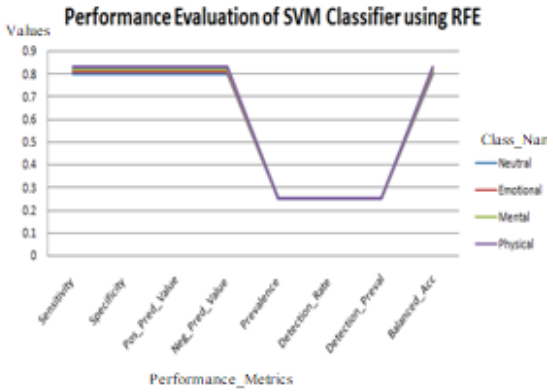

Performance Evaluation of NB Classifier Using RF
Values
Performance_Metrics
Sensitivity
Specificity
Pos_Pred_Value
Neg_Pred_Value
Prevalence
Detection_Rate
Detection_Preval
Balanced_Acc
Neutral
Emotional
Mental
Physical
Target_groups

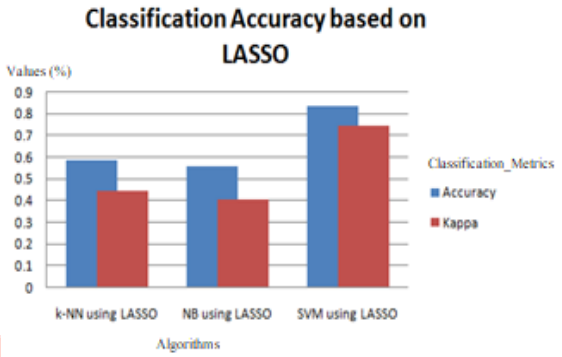

Performance Evaluation of SVM Classifier using RF
Values
Performance_Metrics
Sensitivity
Specificity
Pos_Pred_Value
Neg_Pred_Value
Prevalence
Detection_Rate
Detection_Preval
Balanced_Acc
Neutral
Emotional
Mental
Physical
Target_groups

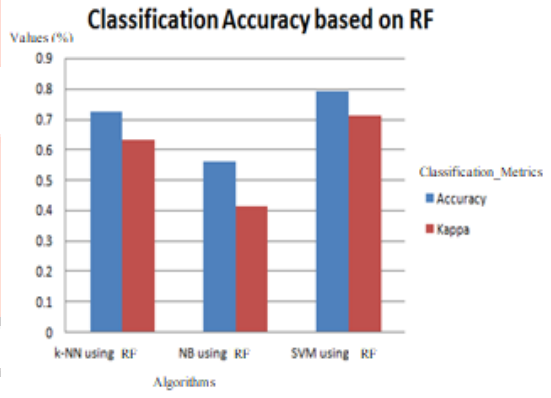

Performance Evaluation of KNN Classifier using RFE
Values
Class_Names
Neutral
Emotional
Mental
Physical
Sensitivity
Specificity
Pos_Pred_Value
Neg_Pred_Value
Prevalence
Detection_Rate
Detection_Preval
Balanced_Acc
Performance_Metrics


Performance Evaluation of NB Classifier using RFE
Values
Class_Names
Neutral
Emotional
Mental
Physical
Sensitivity
Specificity
Pos_Pred_Value
Neg_Pred_Value
Prevalence
Detection_Rate
Detection_Preval
Balanced_Acc
Performance_Metrics


Performance Evaluation of SVM Classifier using RFE
Values
Target_groups
Neutral
Emotional
Mental
Physical
Sensitivity
Specificity
Pos_Pred_Value
Neg_Pred_Value
Prevalence
Detection_Rate
Detection_Preval
Balanced_Acc
Performance_Metrics


Performance Evaluation of SVM Classifier using RFE
Values
Class_Nam
Neutral
Emotional
Mental
Physical
Sensitivity
Specificity
Pos_Pred_Value
Neg_Pred_Value
Prevalence
Detection_Rate
Detection_Preval
Balanced_Acc
Performance_Metrics


Classification Accuracy based on LASSO
Values (%)
Classification_Metrics
Accuracy
Kappa
k-NN using LASSO
NB using LASSO
SVM using LASSO
Algorithms


Classification Accuracy based on RF
Values (%)
Classification_Metrics
Accuracy
Kappa
k-NN using RF
NB using RF
SVM using RF
Algorithms

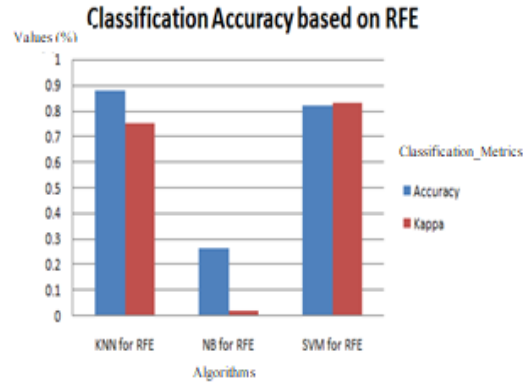

Classification Accuracy based on RFE
Values (%)
Classification_Metrics
Accuracy
Kappa
KNN for RFE
NB for RFE
SVM for RFE
Algorithms

Table 1 Performance Evaluation results based on Accuracy and Kappa

S.No	Name	Accuracy	Kappa
1	k-NN using LASSO	0.58	0.44
2	NB using LASSO	0.55	0.40
3	SVM using LASSO	0.83	0.74
4	KNN using RF	0.72	0.63
5	NB using RF	0.56	0.41
6	SVM using Rf	0.79	0.71
7	KNN using RFE	0.88	0.75
8	NB using RFE	0.26	0.02
9	SVM using RFE	0.82	0.83

5. CONCLUSION AND FUTURE DIRECTION

Better feature identification, which is extremely important in HAR, determines the effectiveness and dependability of a learning process. Existing research mainly focuses on the concept of feature independence to optimize FS algorithms. In other cases, the optimization strategy based on the independence concept leads to poor identification performance. The feature selection techniques and classifiers were critically evaluated in this comparative study. The results for the k-NN classifier using LASSO, RF and RFE, the NB classifier using LASSO, RF and RFE and the SVM classifier using LASSO, RF and RFE are calculated. The charts display sensitivity, specificity, positive prediction value, negative prediction value, prevalence, detection rate, detection prevalence and balanced accuracy. There are four target groups. Physical, emotional, neutral, and mental are the four types. The accuracy and Kappa statistics standard metrics are used to evaluate the forecast. When compared to other methods, SVM + RFE produce the best results shown in Table 1. Various performance metrics are utilized and the consequences are effectively displayed.

In future, filter, wrapper, and embedded methods can all be mixed with FS methods. It's possible that a better outcome will be obtained. This approach should be used on a variety of datasets. For this HAR study, we need better strategies for normalization error. Deep learning is a novel way for this research, therefore rather than using machine learning techniques, Deep learning may provide a superior outcome.

REFERENCES

[1] https://www.linkedin.com/pulse/what-recursive-feature-elimination-amit-mittal/

[2] https://bookdown.org/max/FES/recursive-feature-elimination.html

[3] https://machinelearningmastery.com/rfe-feature-selection-in-python/

[4] https://www.linkedin.com/pulse/what-recursive-feature-elimination-amit-mittal/

[5] https://topepo.github.io/caret/recursive-feature-elimination.html

[6] https://towardsdatascience.com/feature-selection-in-machine-learning-using-lasso-regression-7809c7c2771a

[7] https://medium.com/@23.sargam/lasso-regression-for-feature-selection-8ac2287e25fa

[8] https://corporatefinanceinstitute.com/resources/knowledge/other/lasso/

[9] https://chrisalbon.com/code/machine_learning/trees_and_forests/feature_selection_using_random_forest/

[10] https://blog.datadive.net/selecting-good-features-part-iii-random-forests/

[11] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

[12] https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761

[13] https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

[14] https://www.javatpoint.com/machine-learning-naive-bayes-classifier

[15] https://www.geeksforgeeks.org/support-vector-machine-algorithm/

[16] https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

[17] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[18] geeksforgeeks.org/multi-layer-perceptron-learning-in-tensorflow/

[19] https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron

[20] Proceeding: Thangapriya and Nancy Jasmine Goldena," Comparative analysis of feature selection methods based on activity recognition", published by "Thiruvalluvar university".

[21] Proceeding: Thangapriya and Nancy Jasmine Goldena, "Human Activity Recognition using data mining -A survey", published by "A.P.C Mahalaxmi college for women".