REPUBLIC OF TURKEY

ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

ACTIVITY ANALYSIS AND PREDICTION USING HEALTH DATA

TAHA TURAN AKGÜNGÖR

BACHELOR DEGREE

ADANA 2023

REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER ENGINEERING

ACTIVITY ANALYSIS AND PREDICTION USING HEALTH DATA

TAHA TURAN AKGÜNGÖR
BACHELOR DEGREE

SUPERVISOR
ASST. PROF. DR. MÜMİNE KAYA KELEŞ

ADANA 2023

ABSTRACT

ACTIVITY ANALYSIS AND PREDICTION USING HEALTH DATA

Taha Turan AKGÜNGÖR
Department of Computer Engineering
Supervisor: Asst. Prof. Dr. Mümine KAYA KELEŞ
January 2023, 12 pages

In this study, it is aimed to predict the physical activities of these people by using the health data collected by data mining and taking into account the instant heart rhythm rates of the people. A smart watch was used as a tool for data collection. With the help of this watch, 100 different people were asked to perform 4 different activities, and a total of 10,000 data were collected, 25 from each activity. These 4 activities consist of Sitting, Walking, Stair Climbing and Running.

Keywords: activity analysis, data mining, prediction, health data

ACKNOWLEDGEMENTS

Many people were contacted during the data collection phase of this study. Contributing to the work in line with these communications, especially, I would like to thank my supervisor Asst. Prof. Dr. Mümine Kaya Keleş and I would also like to thank my friends who helped me collect my data.

TABLE OF CONTENTS

TABLE OF CONTENTS iii

LIST OF FIGURES iv

NOMENCLATURE v

1. INTRODUCTION 1

1.1. Literature Review 2

1.2. Data Mining 2

2. MATERIALS AND METHODS 4

2.1. Dataset 4

2.2. WEKA 4

2.3. Machine Learning Algorithms 5

2.3.1. KNN 5

2.3.2. Naive Bayes 6

2.3.3. Random Forest 6

2.3.4. C4.5 6

2.3.5. Support Vector Machine 7

2.3.6. Multi-Layer Perceptron 7

2.3.7. Logistic Regression 7

3. RESULTS AND DISCUSSIONS 8

3.1. Evaluation Criteria of Machine Learning Algorithms 8

3.2. WEKA Results 10

4. CONCLUSIONS 12

REFERENCES 13

LIST OF FIGURES

FIGURE 2.1.1 DATASET AND FEATURES 4
FIGURE 2.2.1 WEKA GUI 5
FIGURE 2.2.2 WEKA EXPLORER 5
FIGURE 3.1.1 LOGISTIC REGRESSION ROC CURVES 8
FIGURE 3.1.2 RANDOM FOREST ROC CURVES 8
FIGURE 3.1.3 MODEL EVALUATIION RESULTS 8
FIGURE 3.1.4 KNN ROC CURVES 9
FIGURE 3.1.5 DECISION TREE ROC CURVES 9
FIGURE 3.1.6 NAÏVE BAYES ROC CURVES 9
FIGURE 3.1.7 MLP ROC CURVES 9
FIGURE 3.1.8 SVM ROC CURVES 9
FIGURE 3.2.1 WEKA MLP CLASSIFIER RESULTS 10
FIGURE 3.2.2 WEKA SVM CLASSIFIER RESULTS 10
FIGURE 3.2.3 SELECTION ATTRIBUTE WITH WEKA 11
FIGURE 3.2.4 ACTIVITY TO HEART RATE GRAPH 11
FIGURE 3.2.5 BMI INDEX TO HEART RATE GRAPH 11

NOMENCLATURE

WEKA	: The Waikato Environment for Knowledge Analysis
KNN	: K-Nearest Neighbors Algorithm
C4.5	: C4.5 is an algorithm used to generate a decision tree.
ROC	: Receiver operating characteristic
RMSE	: Root Mean Square Error

1. INTRODUCTION

In this study, the estimation of the physical activities of these people was carried out by using the health data collected by data mining and taking into account the instantaneous heart rhythm rates of the people. A smart watch was used as a data collection tool. Thanks to this watch, 100 different people were asked to do 4 different activities and a total of 10,000 data were collected, 25 from each activity. These 4 activities consist of Sitting, Walking, Stair Climbing and Running.

Human body activities and activities are an integral part of our lives. Human body activities are our fitness exercises for health management, competitive sport or recovery from injury [1]. Human body activities are for lifting and carrying our child or helping an elderly neighbor carry grocery bag out of the car. The human body in motion runs our daily lives. In doing so, our human body activities contain useful information about our body in motion, whether at work or play, in pain or in joy. [5]

From a business perspective in evaluating our body activities, physiotherapists, trainers, athletes and doctors need quantitative information about human activities. Thanks to the motion estimation study, which was created using health data, this information was estimated.

In this work, the estimation of the physical activities of these people was carried out by using the health data collected by data mining and taking into account the instantaneous heart rhythm rates of the people. A smart watch was used as a data collection tool. Thanks to this watch, 100 different people were asked to do 4 different activities and a total of 10,000 data were collected, 25 from each activity. These 4 activities consist of Sitting, Walking, Stair Climbing and Running. With artificial intelligence, we can see the invisible, measure the immeasurable, and manage human activity as data like any other dataset. In this study, although the number of activities is currently low, the use of artificial intelligence becomes inevitable when a more complex structure is created by using more data over time. Therefore, these data were trained using machine learning algorithms and logical results were obtained.

1.1. Literature Review

In this study, it was investigated which algorithms should be used in the training phase of the data. As a result of these researches, it was decided to use KNN, Naive Bayes, Random Forest, C4.5, Support Vector Machine, Multi-Layer Perceptron and Logistic Regression algorithms.

In order to understand which of these algorithms is more successful, a program called WEKA has been used. As a result of research and studies about WEKA, it has been understood that results such as Classify, Selection of Attributes can be obtained with WEKA. The results obtained with WEKA were compared and it was understood that the most successful algorithm was SVM.

According to literature reviews, 30 participants (referred to as subjects in this dataset) in similar studies performed activities of daily living while carrying a waist-mounted smartphone. The phone is configured to register two applied sensors (accelerometer and gyroscope). For these time series, the administrators of the underlying study performed feature generation and created the dataset by moving a 2.56-second fixed-width window over the series. The result points are equally spaced (1.28s) as the windows overlap 50%. This experiment was videotaped to manually label the data [1].

1.2. Data Mining

With the advancement of technology and being accessible everywhere, many previously physical jobs are now done from places such as computers, mobile phones and tablets. As a result of most transactions with these electronic devices, some data is accumulated on the other side. This data may be meaningless if not processed. Data mining, on the other hand, can be defined as obtaining previously unknown, valid and applicable information from data stacks through a dynamic process.

It is not appropriate to use traditional statistical methods when analyzing big data. Therefore, special methods are needed to process and analyze data. Data mining methods have emerged to meet this need. Although data mining is basically seen as a set of statistical methods, it can be done using mathematical disciplines, modeling techniques, database technology and various computer programs.

There are three commonly used methods in data mining analysis. These methods are; classification, clustering and association rules.

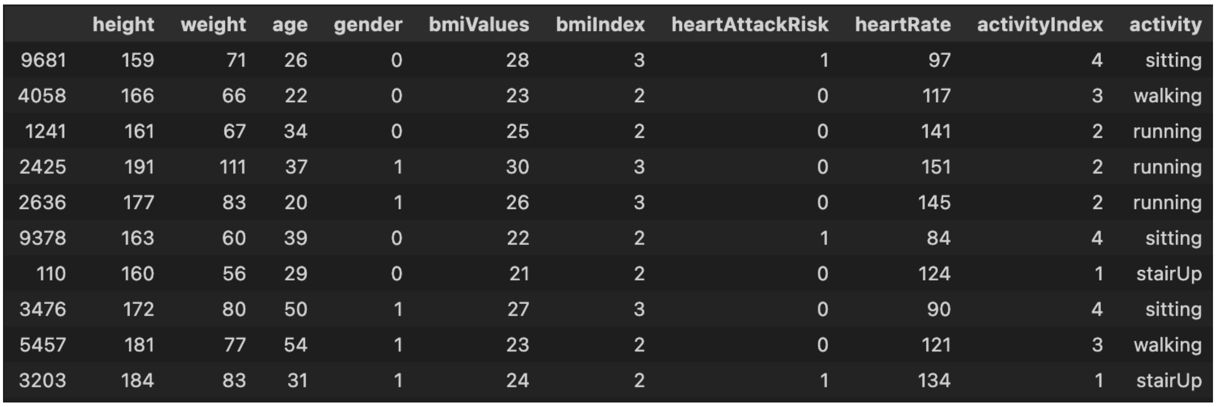
In the classification method, data can be divided into specific or general categories and each can be assigned a class. Practically, decision making processes can be used as a classification problem. With the classification method, a relationship can be established between the values of the data classified and other classified data.

The purpose of the clustering method is to look at the similarities of the data according to their values and to group them accordingly. Similar data are collected in one cluster and different data are collected in another cluster. [3]

2. MATERIALS AND METHODS

2.1. Dataset

The features of the dataset used in this study consist of a total of 9 features: Height, Weight, Age, Sex, Bmi Value, Bmi Index, Heart Attack Risk, Heart Rate, Activity Index. The Activity Index feature is the feature we want to predict. This feature is a section where the activities of Sitting, Walking, Climbing Stairs and Running are digitized. Using these features, a total of 10,000 data were collected, 25 from each activity, for 4 different activities from a total of 100 different people. 70 percent of this data had been used as train data and 30 percent as test data.



	height	weight	age	gender	bmiValues	bmiIndex	heartAttackRisk	heartRate	activityIndex	activity
9681	159	71	26	0	28	3	1	97	4	sitting
4058	166	66	22	0	23	2	0	117	3	walking
1241	161	67	34	0	25	2	0	141	2	running
2425	191	111	37	1	30	3	0	151	2	running
2636	177	83	20	1	26	3	0	145	2	running
9378	163	60	39	0	22	2	1	84	4	sitting
110	160	56	29	0	21	2	0	124	1	stairUp
3476	172	80	50	1	27	3	0	90	4	sitting
5457	181	77	54	1	23	2	0	121	3	walking
3203	184	83	31	1	24	2	1	134	1	stairUp

Figure 2.1.1 Dataset and Features

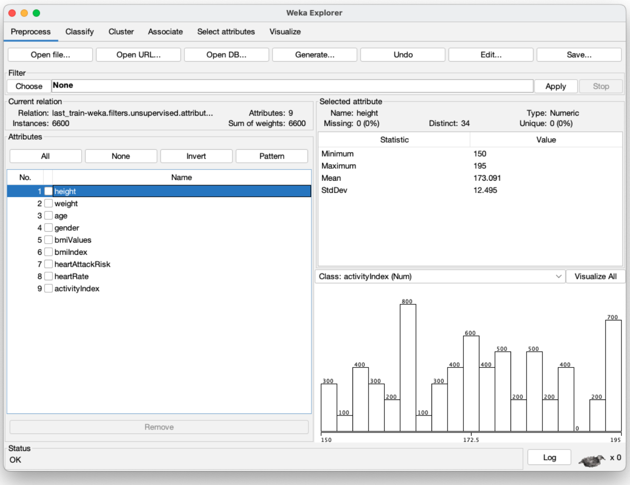
2.2. WEKA

As open-source software, WEKA provides tools for data preprocessing, application of various Machine Learning algorithms, and visualization tools so you can develop machine learning techniques and apply them to real-world data mining problems.

First, it starts with raw data collected from the field. This data may contain several nulls and irrelevant fields. Data preprocessing tools provided in WEKA are used to clean the data. Then, the pre-processed data is saved to local storage to apply machine learning algorithms. Then, depending on the type of machine learning model that is being developed, one of the options such as Classify, Cluster, or Associate is selected. Attribute Selection allows features to be automatically selected to create a reduced dataset.

WEKA provides the implementation of several algorithms under each category. It selects an algorithm selected over the program, sets the desired parameters and is run on the dataset. Next, WEKA gives the statistical output of the model processing. It provides a visualization tool to examine data. Various models can be applied to the same data set. You can then compare the outputs of different models and choose the most suitable for the purpose.

Thus, the use of WEKA results in faster development of machine learning models overall [2]. WEKA interfaces are shown in Figures 2.2.1 and 2.2.2 below.



Filter: None

Current relation
Relation: last_train-weka.filters.unsupervised.attribut... Attributes: 9
Instances: 6600 Sum of weights: 6600

Attributes

No.	Name
1	□ height
2	□ weight
3	□ age
4	□ gender
5	□ bmiValues
6	□ bmiIndex
7	□ heartAttackRisk
8	□ heartRate
9	□ activityIndex

Selected attribute
Name: height Type: Numeric
Missing: 0 (0%) Distinct: 34 Unique: 0 (0%)

Statistic	Value
Minimum	150
Maximum	195
Mean	173.091
StdDev	12.495

Class: activityIndex (Num)

Status
OK

Figure 2.2.2 Weka Explorer

Waikato Environment for Knowledge Analysis
Version 3.8.6
(c) 1999 - 2022
The University of Waikato
Hamilton, New Zealand

Figure 2.2.1 Weka GUI

2.3. Machine Learning Algorithms

A machine learning algorithm is the method by which the AI system conducts its task, generally predicting output values from given input data. The two main processes of machine learning algorithms are classification and regression.

2.3.1. KNN

K-nearest neighbors (KNN) is a type of supervised learning algorithm used for both regression and classification. KNN tries to predict the correct class for the test data by calculating the distance between the test data and all the training points. Then select the K number of points which is closest to the test data. The KNN algorithm calculates the probability of the test data belonging to the classes of 'K' training data and which class holds the highest probability will be selected. In the case of regression, the value is the mean of the 'K' selected training points. [4]

2.3.2. Naive Bayes

A Naive Bayes classifier is a probabilistic machine learning model that's used for classification tasks. The crux of the classifier is based on the Bayes theorem. Theorem is shown in Formula 1 below.

$$P(A\backslash B) = \frac{P(B\backslash A)P(A)}{P(B)} \tag{1}$$

Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is, the presence of one particular feature does not affect the other. Hence it is called naive.

2.3.3. Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

2.3.4. C4.5

The C4.5 algorithm is a famous algorithm in Data Mining. The C4.5 algorithm acts as a Decision Tree Classifier. C4.5 is a data mining algorithm and it is used to generate a decision tree. The C4.5 algorithm is very helpful to generate a useful decision, that is based on a sample of data. C4.5 is given a set of data representing things that are already classified. When we generate the decision trees with the help of C4.5 algorithm, then it can be used for classification of the dataset, and that is the main reason due to which C4.5 is also known as a statistical classifier.

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

2.3.5. Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

2.3.6. Multi-Layer Perceptron

The field of artificial neural networks is often just called neural networks or multi-layer perceptron's after perhaps the most useful type of neural network. A perceptron is a single neuron model that was a precursor to larger neural networks.

It is a field that investigates how simple models of biological brains can be used to solve difficult computational tasks like the predictive modeling tasks we see in machine learning. The goal is not to create realistic models of the brain but instead to develop robust algorithms and data structures that we can use to model difficult problems. The predictive capability of neural networks comes from the hierarchical or multi-layered structure of the networks. The data structure can pick out (learn to represent) features at different scales or resolutions and combine them into higher-order features, for example, from lines to collections of lines to shapes.

2.3.7. Logistic Regression

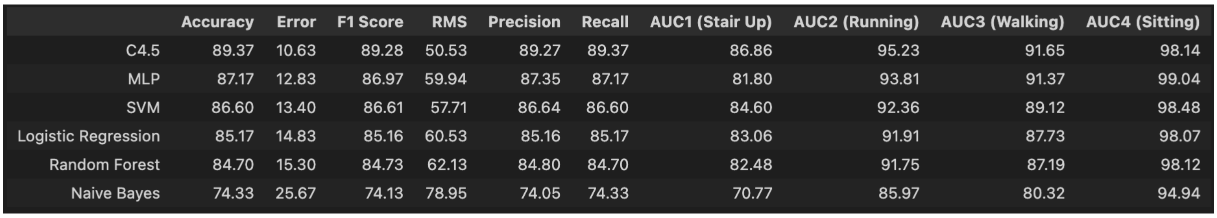
Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on. Multinomial logistic regression can model scenarios where there are more than two possible discrete outcomes. Logistic regression is a useful analysis method for classification problems, where you are trying to determine if a new sample fits best into a category.

3. RESULTS AND DISCUSSIONS

In this thesis, the training of the data and the prediction of the test data with this trained data were made using Machine Learning algorithms. The accuracy rates of these predictions had been found using evaluation criteria and these rates had been compared with all algorithms. According to the results obtained, the most successful algorithm for this dataset was the C4.5 algorithm with an accuracy rate of 89.37 percent.

3.1. Evaluation Criteria of Machine Learning Algorithms

After the model is created using machine learning algorithms, some evaluations are required to understand how these models work. These assessments consist of F-Measure, Accuracy, ROC Area/ROC Curve, RMSE, Precision, and Recall. Using these evaluation criteria, the differences between the algorithms were observed. In the Figures 3.1.1, 3.1.2, 3.1.3, 3.1.4, 3.1.5, 3.1.6, 3.1.7, 3.1.8 below, the evaluation criteria between algorithms are shown as percentages.



	Accuracy	Error	F1 Score	RMS	Precision	Recall	AUC1 (Stair Up)	AUC2 (Running)	AUC3 (Walking)	AUC4 (Sitting)
C4.5	89.37	10.63	89.28	50.53	89.27	89.37	86.86	95.23	91.65	98.14
MLP	87.17	12.83	86.97	59.94	87.35	87.17	81.80	93.81	91.37	99.04
SVM	86.60	13.40	86.61	57.71	86.64	86.60	84.60	92.36	89.12	98.48
Logistic Regression	85.17	14.83	85.16	60.53	85.16	85.17	83.06	91.91	87.73	98.07
Random Forest	84.70	15.30	84.73	62.13	84.80	84.70	82.48	91.75	87.19	98.12
Naive Bayes	74.33	25.67	74.13	78.95	74.05	74.33	70.77	85.97	80.32	94.94

Figure 3.1.3 Model Evaluation Results

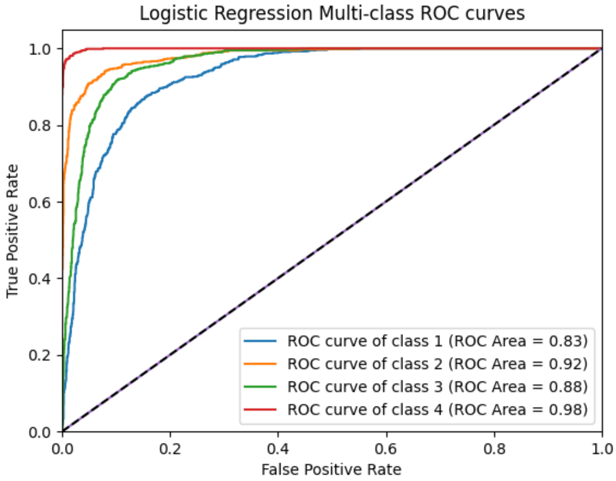


Figure 3.1.1 Logistic Regression ROC Curves

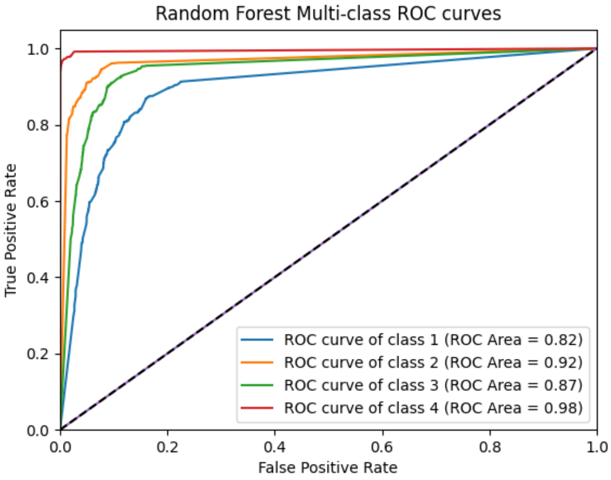


Figure 3.1.2 Random Forest ROC Curves

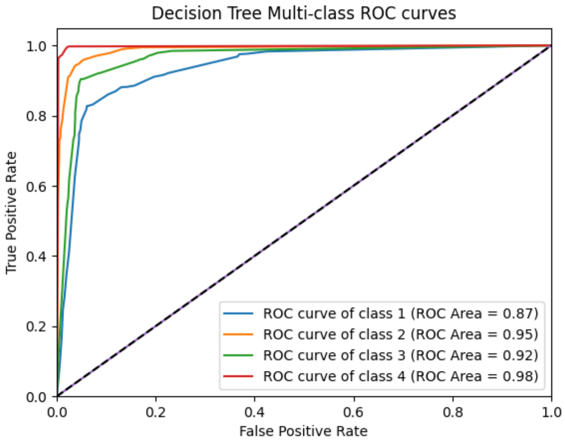


Figure 3.1.5 Decision Tree ROC Curves

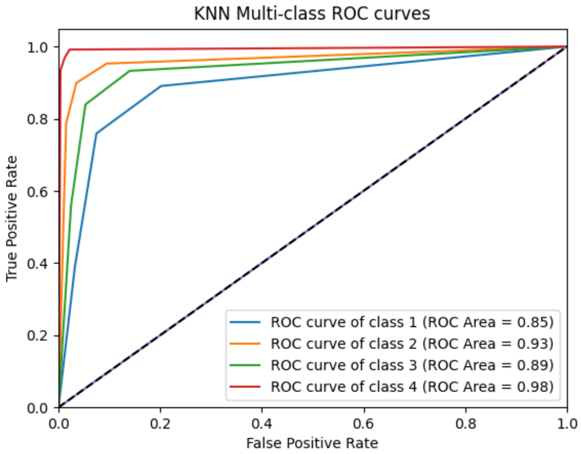


Figure 3.1.4 KNN ROC Curves

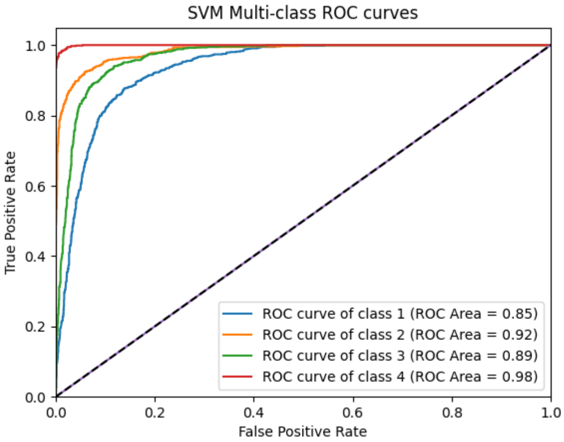


Figure 3.1.8 SVM ROC Curves

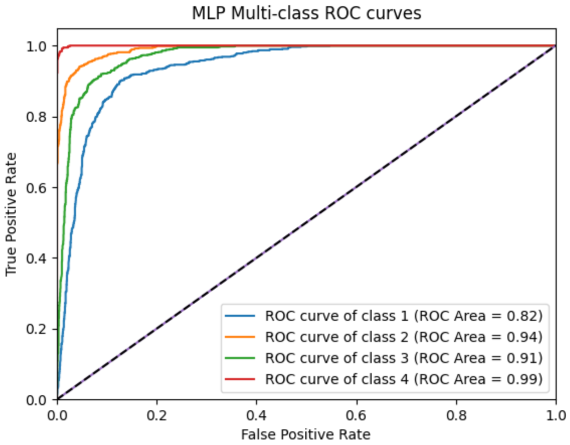


Figure 3.1.7 MLP ROC Curves

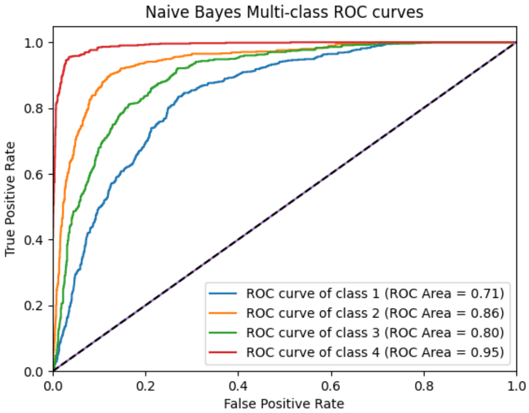
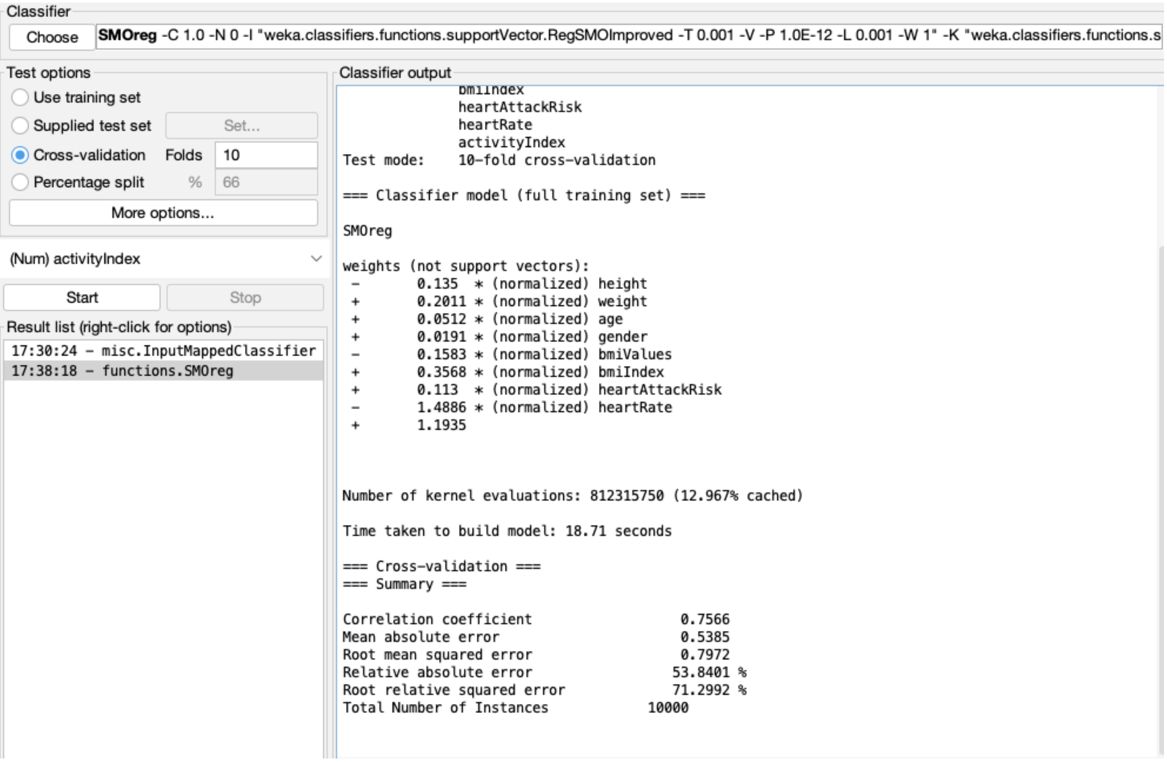


Figure 3.1.6 Naïve Bayes ROC Curves

3.2. WEKA Results

With WEKA, the training results of many machine learning algorithms can be seen. The results of the model trained by WEKA are as in the Figure 3.2.1 and 3.2.2 below.



Classifier: SMOreg -C 1.0 -N 0 -I "weka.classifiers.functions.supportVector.RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "weka.classifiers.functions.s

Test options: ○ Use training set ○ Supplied test set ● Cross-validation Folds 10 ○ Percentage split % 66

(Num) activityIndex

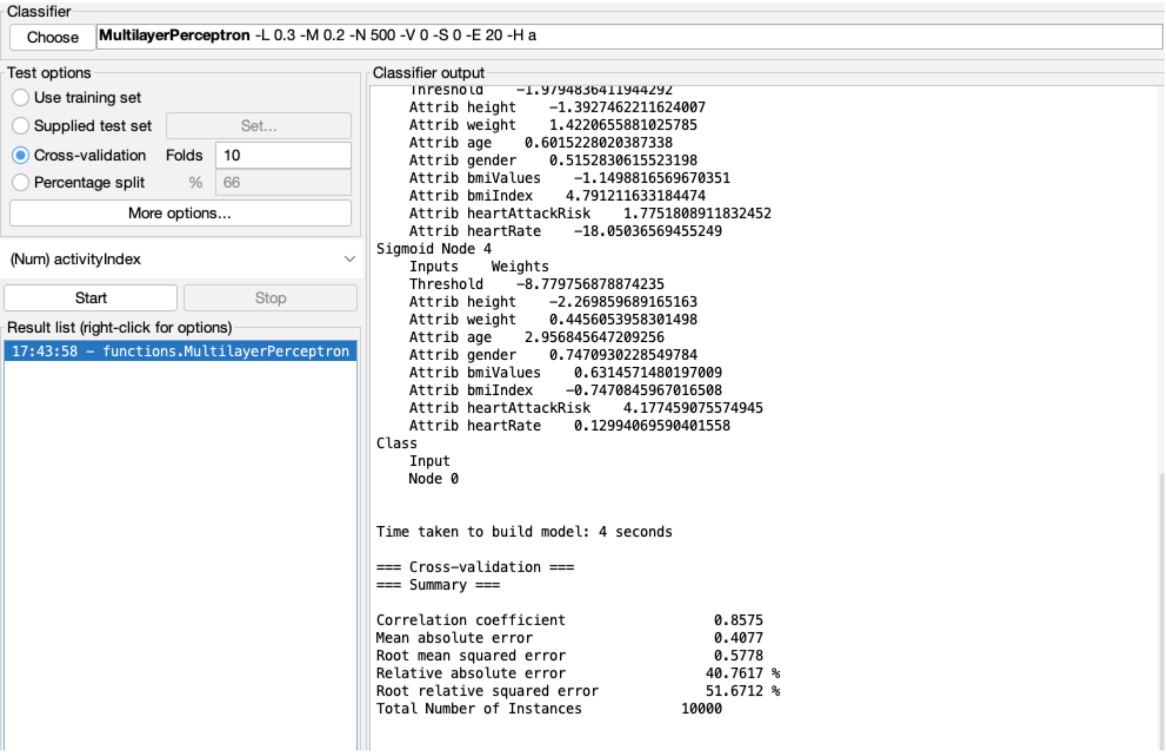
Result list (right-click for options)

```
17:30:24 - misc.InputMappedClassifier
17:38:18 - functions.SMOreg
```

Classifier output

```
              bmiIndex
              heartAttackRisk
              heartRate
              activityIndex
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

SMOreg

weights (not support vectors):
 -       0.135  * (normalized) height
 +       0.2011 * (normalized) weight
 +       0.0512 * (normalized) age
 +       0.0191 * (normalized) gender
 -       0.1583 * (normalized) bmiValues
 +       0.3568 * (normalized) bmiIndex
 +       0.113  * (normalized) heartAttackRisk
 -       1.4886 * (normalized) heartRate
 +       1.1935


Number of kernel evaluations: 812315750 (12.967% cached)

Time taken to build model: 18.71 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.7566
Mean absolute error                      0.5385
Root mean squared error                  0.7972
Relative absolute error                 53.8401 %
Root relative squared error             71.2992 %
Total Number of Instances            10000
```

Figure 3.2.2 Weka SVM Classifier Results



Classifier: MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options: ○ Use training set ○ Supplied test set ● Cross-validation Folds 10 ○ Percentage split % 66

(Num) activityIndex

Result list (right-click for options)

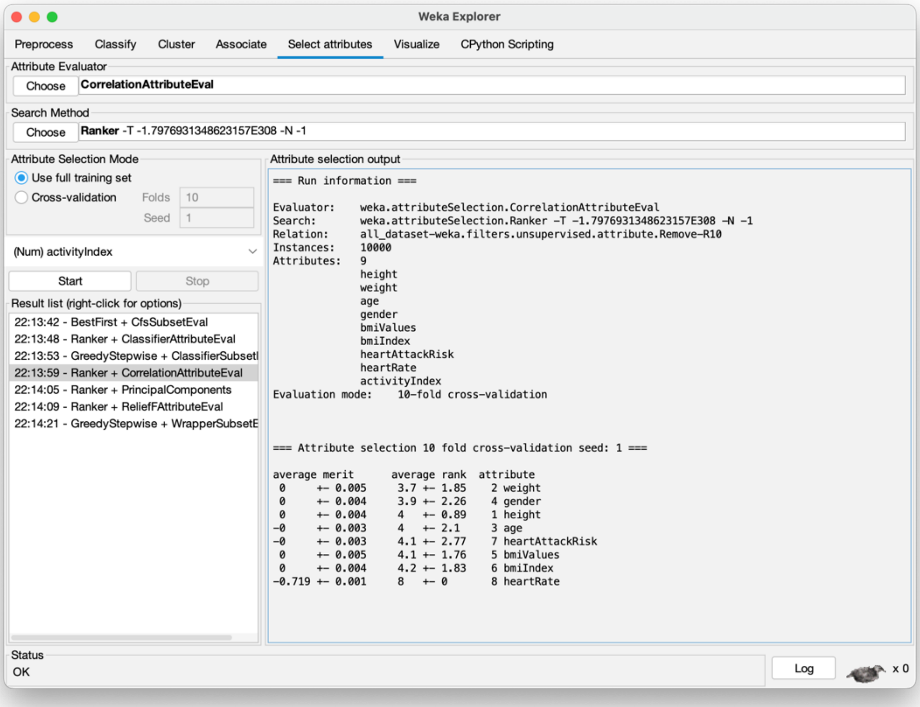
```
17:43:58 - functions.MultilayerPerceptron
```

Classifier output

```
    Threshold    -1.9794836411944292
    Attrib height    -1.3927462211624007
    Attrib weight    1.4220655881025785
    Attrib age    0.6015228020387338
    Attrib gender    0.5152830615523198
    Attrib bmiValues    -1.1498816569670351
    Attrib bmiIndex    4.791211633184474
    Attrib heartAttackRisk    1.7751808911832452
    Attrib heartRate    -18.05036569455249
Sigmoid Node 4
    Inputs    Weights
    Threshold    -8.779756878874235
    Attrib height    -2.269859689165163
    Attrib weight    0.4456053958301498
    Attrib age    2.956845647209256
    Attrib gender    0.7470930228549784
    Attrib bmiValues    0.6314571480197009
    Attrib bmiIndex    -0.7470845967016508
    Attrib heartAttackRisk    4.177459075574945
    Attrib heartRate    0.12994069590401558
Class
    Input
    Node 0


Time taken to build model: 4 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.8575
Mean absolute error                      0.4077
Root mean squared error                  0.5778
Relative absolute error                 40.7617 %
Root relative squared error             51.6712 %
Total Number of Instances            10000
```

Figure 3.2.1 Weka MLP Classifier Results

The results of attribute selection made using Correlation Attribute Evaluation with WEKA are as in the figure below. Attribute selection via WEKA is shown in the Figure 3.2.3 below.



Attribute Evaluator: CorrelationAttributeEval

Search Method: Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode

● Use full training set
○ Cross-validation Folds 10 Seed 1

(Num) activityIndex

Result list (right-click for options)

- 22:13:42 - BestFirst + CfsSubsetEval
- 22:13:48 - Ranker + ClassifierAttributeEval
- 22:13:53 - GreedyStepwise + ClassifierSubsetI
- 22:13:59 - Ranker + CorrelationAttributeEval
- 22:14:05 - Ranker + PrincipalComponents
- 22:14:09 - Ranker + ReliefFAttributeEval
- 22:14:21 - GreedyStepwise + WrapperSubsetE

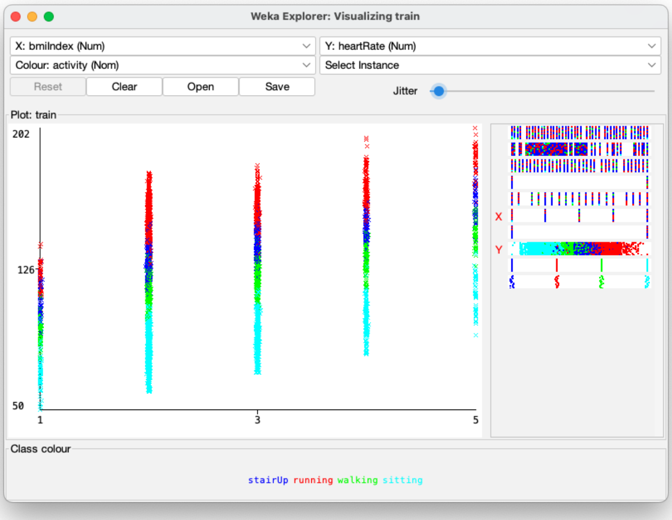
Attribute selection output

```
=== Run information ===

Evaluator:    weka.attributeSelection.CorrelationAttributeEval
Search:       weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:     all_dataset-weka.filters.unsupervised.attribute.Remove-R10
Instances:    10000
Attributes:   9
              height
              weight
              age
              gender
              bmiValues
              bmiIndex
              heartAttackRisk
              heartRate
              activityIndex
Evaluation mode:    10-fold cross-validation



=== Attribute selection 10 fold cross-validation seed: 1 ===

average merit      average rank  attribute
 0     +- 0.005     3.7 +- 1.85    2 weight
 0     +- 0.004     3.9 +- 2.26    4 gender
 0     +- 0.004     4   +- 0.89    1 height
-0     +- 0.003     4   +- 2.1     3 age
-0     +- 0.003     4.1 +- 2.77    7 heartAttackRisk
 0     +- 0.005     4.1 +- 1.76    5 bmiValues
 0     +- 0.004     4.2 +- 1.83    6 bmiIndex
-0.719 +- 0.001     8   +- 0       8 heartRate
```

Status
OK

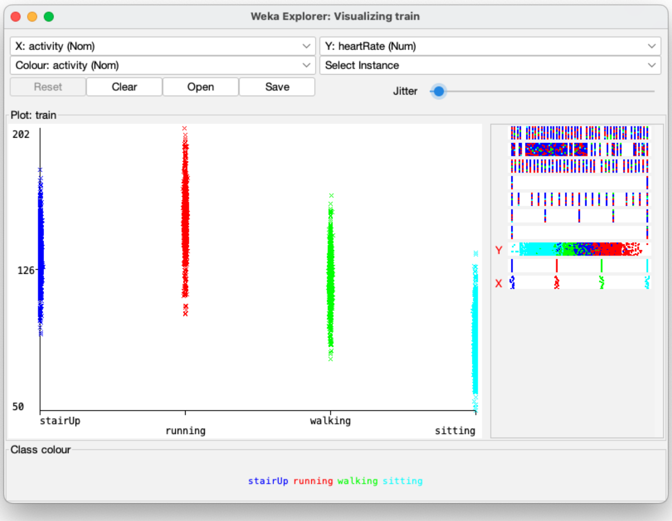
Figure 3.2.3 Selection Attribute with Weka

The graph showing the distribution of heart rate according to body mass index and the distribution of heart rate according to physical activities is as follows. These activities were shown in the Figure 3.2.4 and 3.2.5 below.



X: bmiIndex (Num)
Y: heartRate (Num)
Colour: activity (Nom)
Select Instance
Plot: train
Class colour

Figure 3.2.5 BMI index to Heart Rate Graph



X: activity (Nom)
Y: heartRate (Num)
Colour: activity (Nom)
Select Instance
Plot: train
Class colour

Figure 3.2.4 Activity to Heart Rate Graph

4. CONCLUSIONS

The aim of this thesis is to collect data with experimental methods and to record the collected data in an order, to train these recorded data with certain machine learning algorithms, to obtain a high percentage of accuracy and to test whether these results are really correct. Many proven successful evaluation criteria were used when performing these tests.

In this study, the physical activities of these people were estimated by using the health data collected by data mining and taking into account the instantaneous heart rhythm rates of the people. A smart watch was used as a data collection tool. With the help of this watch, 100 different people were asked to perform 4 different activities and a total of 10,000 data were collected, 25 from each activity. These 4 activities consist of Sitting, Walking, Stair Climbing and Running. The features of the data set used in this study consist of a total of 9 features: Height, Weight, Age, Gender, Bmi Value, Bmi Index, Heart Attack Risk, Heart Rate, Activity Index. The Activity Index property is the one we want to predict.

After the data collection phase was completed, models were created with KNN, Naive Bayes, Random Forest, C4.5, Support Vector Machine, Multi-Layer Perceptron and Logistic Regression machine learning algorithms. After the model was created using machine learning algorithms, some evaluations were made to understand how these models work. These assessments consist of F-Measure, Accuracy, ROC Area/ROC Curve, RMSE, Precision, and Recall. Using these evaluation criteria, differences between algorithms were observed.

Finally, the training results and evaluation results of the models used were obtained by using the WEKA program. These results have been added to the thesis.

In future studies, a model can be created by using the dataset collected from 100 people and the C4.5 algorithm, which is the most successful machine learning algorithm for this dataset. It has been concluded that while estimating with this model, it is very important that the data in the data set are related to each other and that more movement can be predicted if more data is collected for this project.

REFERENCES

[1] Ms.S.Roobini, Ms.J.Fenila Naomi, Smartphone Sensor Based Human Activity Recognition using Deep Learning Models

[2] What is Weka?
https://www.tutorialspoint.com/weka/what_is_weka.htm

[3] Alan, M. A. (2012). Veri madenciliği ve lisansüstü öğrenci verileri üzerine bir uygulama, Dumlupınar University Journal of Social Sciences, 33.

[4] K-Nearest Neighbor
https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4

[5] Your Guide to Human Activities In Machine & Deep Learning
https://medium.com/codex/your-guide-to-human-activities-in-machine-deep-learning-755f59d96295