ACTIVITY ANALYSIS AND PREDICTION USING HEALTH DATA

Taha Turan AKGÜNGÖR
Department of Computer Engineering
Adana, Turkey
180101032@ogr.atu.edu.tr

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.

These data consist of Height, Weight, Age, Gender, Heart Disease Information, Heart Rate and Activity Type. In addition to these features, the body mass index of the person was calculated and these data were added to the csv file as a feature.

	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 1. Dataset and Features

```
Height: 174
Weight: 71
Age: 22
Sex: Male
Body Type / Level: Normal / 2
Body Mass Index: 23
Heart Disease: No
Heart Rate: 120
You are Climbing Stairs now!
```

Figure 2. Output of Model Prediction

Dataset and WEKA results

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. In addition, accuracy rates and attribute selection rates of certain algorithms were obtained through Weka.

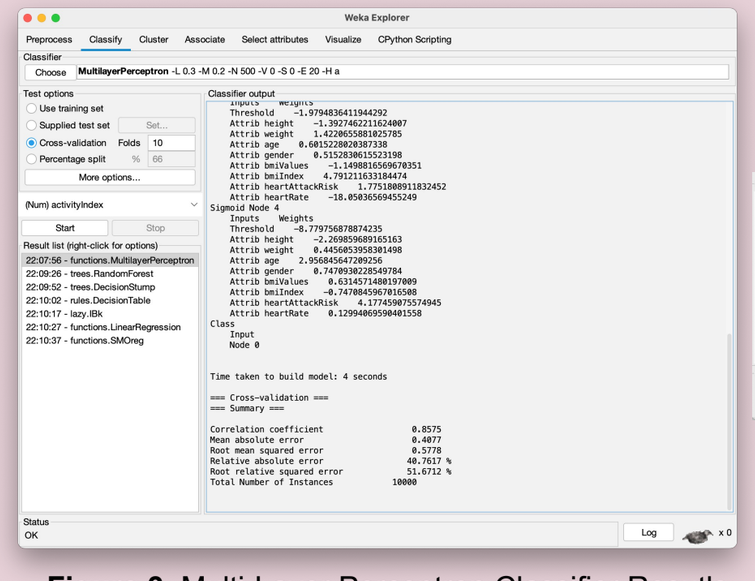

Figure 3. Multi-Layer Perceptron Classifier Resutls

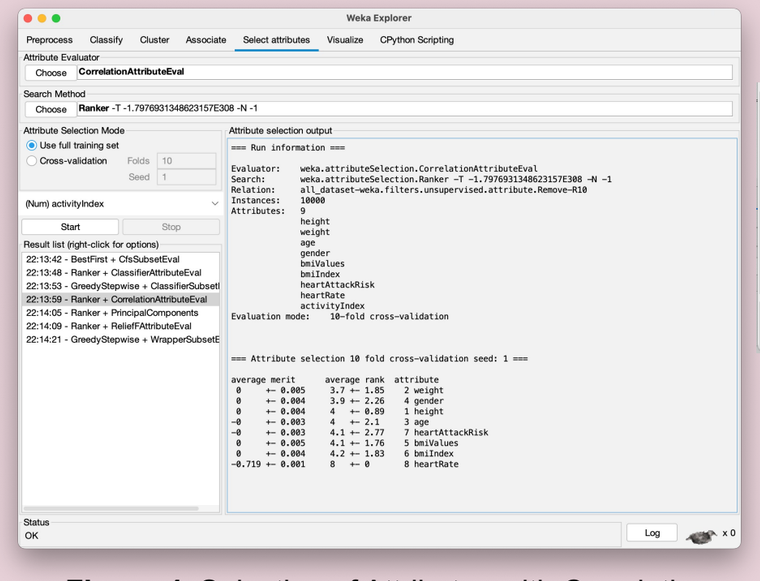

Figure 4. Selection of Attributes with Correlation Attribute Evaluation

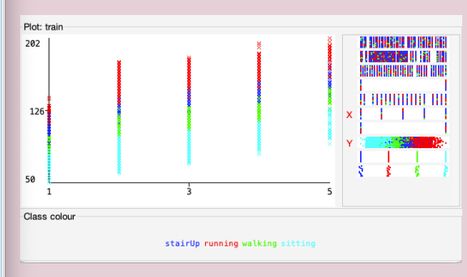

Figure 5. BMI Index to Heart Rate Graph

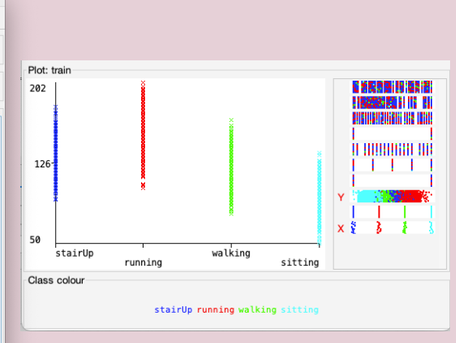

Figure 6. Activity to Heart Rate Graph

Machine Learning Algorithms and Evaluations

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.

	Accuracy	Error	F1 Score	RMS	Precision	Recall
C4.5	89.37	10.63	89.28	50.53	89.27	89.37
MLP	87.17	12.83	86.97	59.94	87.35	87.17
SVM	86.60	13.40	86.61	57.71	86.64	86.60
Logistic Regression	85.17	14.83	85.16	60.53	85.16	85.17
Random Forest	84.70	15.30	84.73	62.13	84.80	84.70
Naive Bayes	74.33	25.67	74.13	78.95	74.05	74.33

Figure 7. Evaluation Criteria results

AUC1 (Stair Up)	AUC2 (Running)	AUC3 (Walking)	AUC4 (Sitting)
86.86	95.23	91.65	98.14
81.80	93.81	91.37	99.04
84.60	92.36	89.12	98.48
83.06	91.91	87.73	98.07
82.48	91.75	87.19	98.12
70.77	85.97	80.32	94.94

Figure 8. ROC-AUC results

Results

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.

Conclusion

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