New Trends and Issues Proceedings on Humanities and Social Sciences

Volume 5,Issue 4 (2018) 58-64

ISSN 2547-8818
www.prosoc.eu

Selected Paper of 7th Cyprus International Conference on Educational Research (CYICER-2018) 07-09 June, 2018, Acapulco Hotel Convention Center in Kyrenia, Cyprus

Multiple linear regression-based physical fitness prediction models for Turkish secondary school students

M. Fatih Akay*, Department of Computer Engineering, Cukurova University, 01330 Adana, Turkey
Ozge Bozkurt, Department of Computer Engineering, Cukurova University, 01330 Adana, Turkey
Ebru Cetin, School of Physical Education and Sport, Gazi University, 06560 Ankara, Turkey
Imdat Yarim, School of Physical Education and Sport, Gazi University, 06560 Ankara, Turkey

Suggested Citation:
Fatih Akay, M., Bozkurt, O., Cetin, E. & Yarim, I. (2018). Multiple linear regression-based physical fitness prediction models for Turkish secondary school students. *New Trends and Issues Proceedings on Humanities and Social Sciences.* [Online]. *5*(4), pp 58–64. Available from: www.prosoc.eu

Selection and peer review under responsibility of Prof. Dr. Huseyin Uzunboylu, Near East University, Cyprus

Abstract

Physical fitness is a necessary component for daily activities. Measurement of physical activity is essential for determining physical fitness rate. This study aims to develop new prediction models for predicting the physical fitness of Turkish secondary school students by using multiple linear regression (MLR). The datasets comprise data of various number of subjects according to the target variables including the test scores of the 30m speed, 20m stage run, balance and hand-grip (right/left). The predictor variables used to develop the prediction models are gender, age, body mass index (BMI), body fat, number of curl-up and push-ups in 30 seconds. Eight physical fitness prediction models for each target have been created with the predictor variables listed above. The performance of the prediction models has been calculated by using standard error of estimate (SEE). The results show that MLR-based prediction models can be safely used to predict the physical fitness of Turkish secondary school students.

Keywords: Physical fitness, multiple linear regression, machine learning, validation.

* ADDRESS FOR CORRESPONDENCE: **M. Fatih Akay**, Department of Computer Engineering, Cukurova University, 01330 Adana, Turkey. *E-mail address:* mfakay@cu.edu.tr / Tel.: +90-322-338-7101

1. Introduction

Physical fitness is defined as any bodily movement produced by skeletal muscles that result in energy expenditure. Physical fitness is a necessary component for daily activities, which keeps your body and mind healthy and it is important for maintaining healthy lifestyles. To successfully detect health level, measurement of physical fitness has become a fundamental component for health monitoring as part of public health measures. Performance or skill-related fitness typically focusses on improving agility, balance, speed, power, coordination and reaction time (Witten, Frank, Hall & Pal, 2016).

Physical tests have a significant role to define links between fitness, health and physical activity. Fitness training and physical fitness tests provide assessment, monitoring and evaluation opportunities but measurement of the physical activity to achieve the physical fitness rate needs professional equipment, experienced staff and a lot of time. Because of these deficiencies of direct measurement, several physical fitness prediction models using statistical and machine learning methods have been proposed in the literature. In Jaakkola, Yli-Piipari, Huotari, Watt & Liukkonen (2016), the purpose of this study was to examine the extent to which fundamental movement skills and physical fitness scores assessed in early adolescence predict self-reported physical activity assessed 6 years later. The sample comprised of 333 (200 girls and 133 boys) students. The effects of previous physical activity, sex and BMI were investigated in the main analysis. Adolescents' fundamental movement skills, physical fitness, self-report physical activity and BMI were collected at baseline, and their self-report energy expenditure (metabolic equivalents: METs) and intensity of physical activity were collected using the International physical activity questionnaire 6 years later. Results showed that after controlling for previous levels of physical activity, sex and BMI, the size of the effect of fundamental movement skills and physical fitness on energy expenditure and physical activity intensity was moderate, with the effect being stronger for high-intensity physical activity. In Fergus et al. (2015), a supervised machine learning approach was used for measuring and monitoring physical activity for children. A supervised machine learning approach has been adopted by a set of activities and features suitable for measuring physical activity and evaluates the use of a multilayer perceptron neural network (NN) for the number of 28 subjects. The paper presents a set of activities and features to measure physical activity and assesses the use of a multi-layered predictive NN to classify physical activities by activity type. Results showed the possibility to obtain an overall accuracy when three and four feature combinations were used. In Ahmed & Loutfi (2013) case-based reasoning (CBR) approach to identify the physical activity of elderly based on pulse rate has been proposed. The CBR approach has been compared with the two popular classification techniques including SVM and NN on 24 subjects. In Dijkhuis, Blaauw, Van Ittersum & Aiello (2018) an activity tracker to record participants' daily step count has been used as input for a coaching session. The gathered step count data were used to train eight different machine learning algorithms to make hourly estimations of the probability of achieving a personalised, daily steps threshold for 48 subjects. In Reichherzer, Timm, Earley & Reyes (2017) the data analysis methods have been used to train a classifier for records with the individuals, their physical activities and conditions under which they were performed. Four different machine learning algorithms including Decision Tree, Random Forest, SVM and Naive Bayesian were used to make predictions for 29 participants. Results of the study showed that the methods allow individuals and their activities to be tracked.

As we can observe from literature, all the studies concentrate on classifying physical activity levels rather than predicting the actual test results. The developed models are not suitable for classifying the physical activity level of Turkish children. The developed models require the subject to complete several exercises in order to have his/her physical activity level classified. The aim of this study is to develop new prediction models for Turkish secondary school students by using MLR.

The content of the paper is structured as follows. Section 2 describes dataset generation. Section 3 presents results and discussion. Lastly, Section 4 concludes the paper.

2. Dataset generation

The dataset comprises of different number of subjects depending on the target variables of healthy secondary school students. Subjects' data used in this study were taken from the College of Physical Education and Sports at Gazi University. Physical exercise tests were applied to those subjects to acquire their physical fitness. Before the physical exercise tests, a consent participant form was signed by all subjects participating in this study. Participants were assigned to perform the following core stabilisation assessments 30 meter (m) speed, 20m stage run, hand grip (right/left) and balance tests.

Statistical information about the datasets is given in Table 1– 5.

Table 1. Statistics of the 30m speed test

Variables	Minimum	Maximum	Mean	Standard deviation
Gender	–	–	–	–
Age (year)	11	16	12.75	0.9
Height (cm)	130	182	152.99	8.45
Weight (kg)	46.6	107.6	46.509	12.41
BMI (kg/m^2)	13.5	43.7	19.82	3.97
Curl-up	0	26	11.88	5.07
Push-up	0	30	8.86	7.04
Body fat (%)	9.6	41.7	23.03	6.54
30m speed (s)	4.4	8.8	5.98	0.6

Table 2. Statistics of the 20m stage run test

Variables	Minimum	Maximum	Mean	Standard deviation
Gender	–	–	–	–
Age (year)	11	16	12.81	0.9
Height (cm)	130	182	153.11	12.09
Weight (kg)	46.6	107.6	46.87	12.47
BMI (kg/m2)	14	32.5	19.79	3.72
Curl-up	0	26	11.89	5.12
Push-up	0	30	8.99	7.03
Body fat (%)	9.6	41.7	23.02	6.55
20m stage run (s)	2	32.1	4.35	2.21

Table 3. Statistics of the hand-grip (right) test

Variables	Minimum	Maximum	Mean	Standard deviation
Gender	–	–	–	–
Age (year)	11	16	12.75	0.9
Height (cm)	130	182	152.99	8.44
Weight (kg)	46.5	107.6	46.63	12.23
BMI (kg/m2)	13.5	43.7	19.81	3.97
Curl-up	0	26	11.86	5.07
Push-up	0	30	8.83	7.05
Body fat (%)	9.6	41.7	23.02	6.53
Hand-grip (cm)	10.1	43.1	22.54	5.43

Table 4. Statistics of the hand-grip (left) test

Variables	Minimum	Maximum	Mean	Standard deviation
Gender	–	–	–	–
Age (year)	11	16	12.75	0.9
Height (cm)	130	182	152.99	8.44
Weight (kg)	46.5	107.6	46.63	12.70
BMI (kg/m2)	13.5	43.7	19.81	3.97
Curl-up	0	26	11.86	5.07
Push-up	0	30	8.83	7.05
Body fat (%)	6.53	41.7	23.03	6.65
Hand-grip (cm)	10.2	40.6	21.94	5.32

Table 5. Statistics of the balance test dataset

Variables	Minimum	Maximum	Mean	Standard deviation
Gender	–	–	–	–
Age (year)	11	16	12.75	0.9
Height (cm)	130	182	153.1	8.56
Weight (kg)	46.5	107.6	46.44	12.48
BMI (kg/m2)	14	43.7	19.77	3.96
Curl-up	0	26	11.82	5.08
Push-up	0	30	9	7.02
Body fat (%)	9.6	41.7	22.94	6.53
Balance (s)	0	73.07	6.33	7.02

3. Results and discussion

MLR is employed in order to develop physical fitness prediction models. MLR is the most common form of linear regression analysis. As a predictive analysis, MLR is used to explain the relationship between one continuous dependent variable and two or more independent variables. The variable that we want to predict is called the dependent variable or target variable. The independent variables can be continuous (age, weight and height) or categorical (gender). MLR analysis helps to understand how much will the dependent variable change when we change the independent variables.

By using combinations of the predictor variables, eight different prediction models have been produced. The performance of the prediction models has been evaluated using SEE, the equation of which is given in Eq. (1).

$$SEE = \sqrt{\frac{\sum (Y - Y')^2}{N}} \tag{1}$$

In Eq. (1), *Y* is the measured value, *Y*' is the predicted value and *N* is the number of samples in a test subset.

Table 6 shows the predictor variables of physical fitness prediction models. Tables 7–11 show SEE's of physical fitness prediction models.

Table 6. Combinations of predictor variables

Model nos.	Predictor variables
1	Gender, age, BMI
2	Gender, age, BMI, curl-up
3	Gender, age, BMI, push-up
4	Gender, age, BMI, body fat
5	Gender, age, BMI, curl-up, push-up
6	Gender, age, BMI, curl-up, body fat
7	Gender, age, BMI, push-up, body fat
8	Gender, age, BMI, curl-up, push-up, body fat

Table 7. SEE values for 30m speed test prediction models

Model nos.	No validation	10-fold	5-fold	Random 80–20%	Random 70–30%
1	0.51	0.52	0.52	0.58	0.57
2	0.46	0.47	0.46	0.53	0.50
3	0.48	0.49	0.49	0.53	0.54
4	0.48	0.50	0.49	0.53	0.55
5	0.44	0.45	0.45	0.51	0.49
6	0.43	0.46	0.45	0.49	0.49
7	0.46	0.48	0.48	0.50	0.53
8	0.42	0.45	0.44	0.47	0.48

Table 8. SEE values for 20m stage run prediction models

Model nos.	No validation	10-fold	5-fold	Random 80–20%	Random 70–30%
1	2.08	2.10	2.09	3.92	3.28
2	2.05	2.07	2.06	3.90	3.25
3	2.05	2.08	2.07	3.93	3.30
4	2.08	2.11	2.11	3.95	3.33
5	2.03	2.06	2.05	3.92	3.27
6	2.05	2.09	2.08	3.94	3.31
7	2.05	2.09	2.09	3.96	3.34
8	2.03	2.08	2.07	3.95	3.32

Table 9. SEE values for hand grip (right) test prediction models

Model nos.	No validation	10-fold	5-fold	Random 80–20%	Random 70–30%
1	4.64	4.68	4.69	5.39	4.79
2	4.56	4.61	4.62	5.38	4.83
3	4.60	4.67	4.66	5.45	4.91
4	4.61	4.81	4.81	6.36	4.76
5	4.54	4.61	4.61	5.45	4.91
6	4.54	4.76	4.76	6.25	4.81
7	4.59	4.81	4.78	6.28	4.88
8	4.53	4.76	4.75	6.20	4.89

Table 10. SEE values for hand grip (left) test prediction models

Model nos.	No validation	10-fold	5-fold	Random 80–20%	Random 70–30%
1	4.37	4.44	4.39	4.51	4.49
2	4.32	4.40	4.35	4.39	4.40
3	4.32	4.39	4.38	4.47	4.44
4	4.30	4.78	4.75	4.35	4.38
5	4.28	4.38	4.36	4.39	4.38
6	4.26	4.75	4.73	4.33	4.35
7	4.27	4.75	4.77	4.29	4.33
8	4.24	4.74	4.75	4.29	4.32

Table 11. SEE values for balance test prediction models

Model nos.	No validation	10-fold	5-fold	Random 80–20%	Random 70–30%
1	6.60	6.66	6.67	9.39	8.22
2	6.57	6.65	6.67	9.34	8.21
3	6.46	6.54	6.56	9.15	8.00
4	6.59	6.65	6.66	9.37	8.34
5	6.45	6.55	6.56	9.14	8.04
6	6.57	6.65	6.66	9.32	8.36
7	6.46	6.55	6.57	9.16	8.14
8	6.45	6.55	6.57	9.14	8.19

Among the physical fitness tests, the best prediction performance has been obtained for 30m speed test. On the other hand, the worst prediction performance has been obtained for balance test. As far as the results are concerned, no significant differences among the variables have been observed. In terms of validation features, based on our observations from the obtained results, the 'No Validation' produced the best prediction results whereas 'Random 80–20%' division yielded the worst prediction performance.

4. Conclusion

In this study, eight different physical fitness prediction models have been developed by using MLR for Turkish secondary school students. The results show that MLR is a viable method for physical fitness prediction with acceptable *SEE* values.

Future work can be performed by using machine learning methods combined with feature selection algorithms to improve the accuracy of physical fitness prediction.

Acknowledgements

The authors would like to thank Cukurova University Scientific Research Projects Center for supporting this work under Grant no. FYL-2018-10430.

References

Ahmed, M. U. & Loutfi, A. (2013). Pyhsical activity identification using supervised machine learning and based on pulse rate. *International Journal of Computer Science and Applications, 4*(7).

Dijkhuis, T. B., Blaauw, F., Van Ittersum, M. W. & Aiello, M. (2018). Personalized physical activity coaching, a machine learning approach. *Sensors, 18*(2), 623.

Fergus, P., Hussain, A., Hearty, J., Fairclough, S., Boddy, L., .·.·.· Radi, N. (2015). *A machine learning approach to measure and monitor physical activity in children to help fight overweight and obesity*. In 11th International Conference on Intelligent Computing Theories and Methodologies, ICIC 2015 (pp. 676–688). Fuzhou, China.

Jaakkola, T., Yli-Piipari, S., Huotari, P., Watt, A. & Liukkonen, J. (2016). Fundamental movement skills and physical fitness as predictors of physical activity, A 6-year follow-up study. *Scandinavian Journal of Medicine & Science in sports, 26*(1), 74–81.

Reichherzer, T., Timm, M., Earley, N. & Reyes, N. (2017). Using machine learning techniques to track individuals & their fitness activities. In A. Bossard, G. Lee & L. Miller (Eds.), *Proceedings of 32nd International Conference on Computers and Their Applications*, Honolulu, Hawaii, USA (pp. 119–124). Winona, MN: ISCA.

Witten, I. H, Frank, E., Hall, M. A. & Pal, C. J. (2016). *Data mining: practical machine learning tools and techniques, Morgan Kaufmann series in data management systems* (4th ed.).