**Run or Walk Analysis**

PORNIMA BANSODE

Dept. of Information Systems, Spring 2018

Northeastern University, Boston, MA

bansode.p@husky.neu.edu

**Abstract:**

Walking and running are the most popular physical activities for adults. There is always a debate on which is more preferable, running or walking. The research suggests that for weight loss, running is the best way compared to walking. Few studies also show that runners are thinner than the walkers. It is found that a person weighing 160 pounds can burn over 800 calories per hour by running 8 mph compared to a person burning 300 calories at 3.5 mph by walking.

**Introduction:**

We aim at building different predictive models with high accuracy which will enable us to analyse people who are running or walking on the street. The dataset consists of 88588 sensor data samples which is collected from accelerometer and gyroscope in 10seconds interval and 5.4/second frequency. The data comprises columns acceleration\_x, acceleration\_y, acceleration\_z, gyro\_x, gyro\_y, gyro\_z.

There is an activity type represented by "activity" column which acts as label and reflects following activities:

"0": walking and "1": running

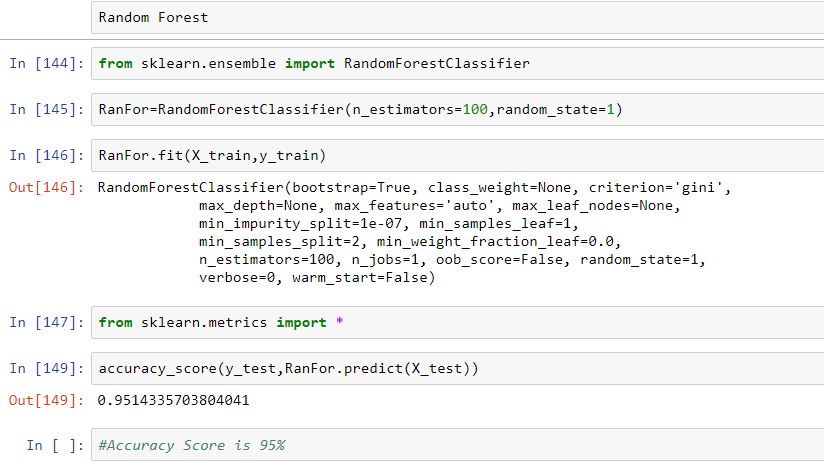
Also, the dataset contains "wrist" column which represents the wrist where the device was placed to collect a sample on:

"0": left wrist and "1": right wrist

Additionally, the dataset contains "date", "time" and "username" columns which provide information about the exact date, time and user which collected these measurements.

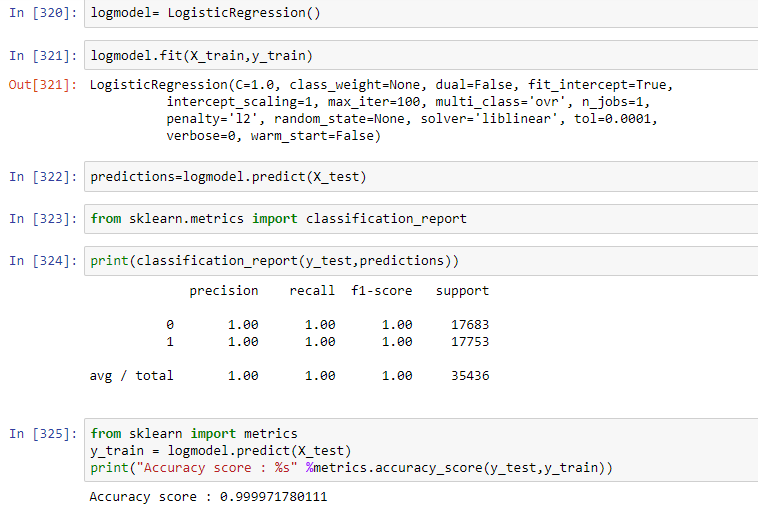
**Code with Documentation:**

**Random Forest**



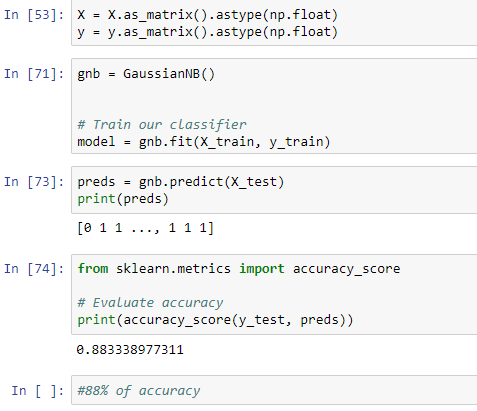
Random Forest gave the accuracy of 95%.

**Logistic Regression**



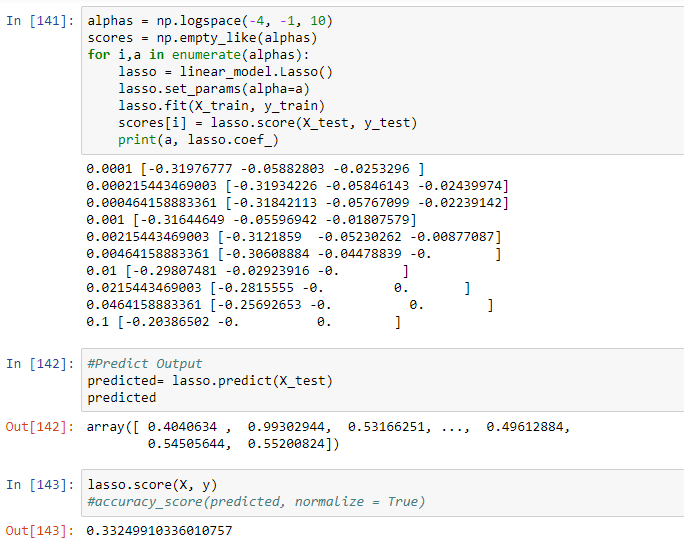
Logistic Regression gave accuracy of 99%.

**Naïve Bayes Classification**



Accuracy of 88%

**Lasso Regression**



Lasso Regression gave accuracy of 33%

**Results:**

From the training results, we analysed that after applying Logistic Regression model with accuracy score of 0.9999 which proves to be the optimum model. The next model is the Random Forest the accuracy score was 0.9514. Also, Naïve Bayes Classification model can be considered as a good model with accuracy of 0.8833.

**Discussion:**

Initially we applied the existed model to our dataset and analysed which model gives the best precision and prediction. The dataset consists of various float values, also the dependent variables are related to each other. The plotting shows there is a significant linear relationship between the variables. Hence, we considered the linear model here. Further, to analyse which model fits the best, we tried to use different predictive models to get the highest accuracy score. Our aim was to get the optimum model for the present dataset. Logistic Regresion was found to be the best model with highest accuracy. To get more knowledge on the dataset we can work with TensorFlow that may give more precise accuracy.

**Data Source:**

<https://www.kaggle.com/vmalyi/run-or-walk>

# References

*Building Random Forest*. (n.d.). Retrieved from http://dataaspirant.com/2017/06/26/random-forest-classifier-python-scikit-learn/

*Decision Trees*. (n.d.). Retrieved from https://www.packtpub.com/books/content/predicting-sports-winners-decision-trees-and-pandas

GOLDE, K. (2015, 07 06). Retrieved from Gretist: https://greatist.com/fitness/walking-good-workout-running

*In Depth: Naive Bayes Classification*. (n.d.). Retrieved from https://jakevdp.github.io/PythonDataScienceHandbook/05.05-naive-bayes.html

*Lasso Regression*. (n.d.). Retrieved from https://chrisalbon.com/machine-learning/lasso\_regression.html

Polamuri, S. (2017, 26 06). *BUILDING RANDOM FOREST CLASSIFIER WITH PYTHON SCIKIT LEARN*. Retrieved from dataaspirant: http://dataaspirant.com/2017/06/26/random-forest-classifier-python-scikit-learn/

REYNOLDS, G. (2013, 05 29). *Is it better to walk or run?* Retrieved from https://well.blogs.nytimes.com/2013/05/29/is-it-better-to-walk-or-run/