**BODY\_FITNESS­ REPORT**

**Introduction:**

The motivation behind collecting this data-set is personal for Ravi. Ravi noticed for a while that during the months when he is exercising regularly, he felt more active and would move around a lot more, as opposed to when he was not working out, he would feel lethargic. [Ravi wanted to know the connection between exercise, stepcount and activeness.]

**Objective:**

1. To investigate whether exercise/ working-out will increase daily step-count
2. To examine the association between exercise/ working-out and person’s activeness

**Data Dictionary:**

* Date
* Step\_Count
* Mood
* Calories\_Burned
* Hours\_Of\_Sleep
* Bool\_Of\_Active
* Weight\_Kg

**Advantages of Exercise:**

* It can make you feel happier:
* It can help with weight loss
* It is good for your muscles and bones
* it can increase your energy levels
* It can reduce your risk of chronic disease
* it can help skin health
* It can help your brain health and memory
* It can help your brain health and memory
* it can help with relaxation and sleep quality
* It can reduce pain
* it can promote a better sex life

**Disadvantages of Exercise:**

* Exercise is addictive.
* Exercise hurts the heart
* exercise is associated with body perception disorders.
* exercise can break up families.
* Exercise can cause diabetes
* Exercise destroys diets
* Exercise causes inflammation.
* exercise is stressful.

**Proposed study**

* Objective 1 - Multiple Linear Regression
* Objective 2 - Naive Bayes

**Multiple Linear Regression:**

If there are two or more independent variables and one dependent variable then multiple linear regression is used. Multiple linear regression is a supervised learning. It is statistical tool used for modeling and establishes a relationship between the one or more independent data(x) and dependent data(y) .

The representation is a linear equation that combines a specific set of input values(x) the solution to which is the predicted output for that set of input values(y). as such, both the input(x) and output values are numeric.

**Naive Bayes:**

 naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem  with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models. But they could be coupled with Kernel density estimation and achieve higher accuracy levels.

**Research questions:-**

**Question– 1** Does exercise/ working-out increase daily step-count?

**Hypothesis**

 H0: There is no association between exercise and increase in daily step-count

H1: There is a significant association between exercise and increase in daily step-count

Dependent Feature/ Variable/ Attribute

* stepcount

Independent Features/ Variables/ Atributes

* caloriesburnt
* sleephrs
* weight
* mood\_Happy, mood\_Neutral, mood\_Sad
* activeness\_Active, activeness\_In-active

**Question - 2** Does exercise/ working-out improve a person’s activeness?

**Hypothesis**

H0: There is no association between exercise and improvement in person’s activeness

H1: There is a significant association between exercise and improvement in person’s activeness

Dependent Feature/ Variable/ Attribute

* activeness

Independent Features/ Variables/ Atributes

* stepcount
* caloriesburnt
* sleephrs
* weight
* mood

**Interpretations:**

We have two research objectives in this project. In our first objective we have analysed the association between step count as the dependent variable and calories burned ,sleep hours, weight mood Happy, mood Neutral, mood Sad ,activeness Active and activeness In-active as the independent variables. We have applied simple linear regression here

**This is the results we got from linear regression**

""" Scientific Notation - 1.2e-6 = 0.0000012 5.48E+03 = 5482.52291

y = -2761.42260035 + (29.36748323 \* caloriesburnt) + (-2.39621345 \* sleephrs)

+ (46.13251252 \* weight) + (24.27763435 \* mood\_Happy)

+ (66.28802792 \* mood\_Neutral) + (-90.56566227 \* mood\_Sad)

+ (10.56934685 \* activeness\_Active) + (-10.56934685 \* activeness\_In-active)

* Holding all other features fixed, a 1 unit increase in caloriesburnt is associated with an increase in stepcount by 29
* Holding all other features fixed, a 1 unit increase in sleephrs is associated with an decrease in stepcount by 2
* Holding all other features fixed, a 1 unit increase in weight is associated with an increase in stepcount by 46
* Holding all other features fixed, a 1 unit increase in mood\_Happy is associated with an increase in stepcount by 24
* Holding all other features fixed, a 1 unit increase in mood\_Neutral is associated with an increase in stepcount by 66
* Holding all other features fixed, a 1 unit increase in mood\_Sad is associated with a decrease in stepcount by 91
* Holding all other features fixed, a 1 unit increase in activeness\_Active is associated with an increase in stepcount by 11
* Holding all other features fixed, a 1 unit increase in activeness\_In-active is associated with a decrease in stepcount by 11

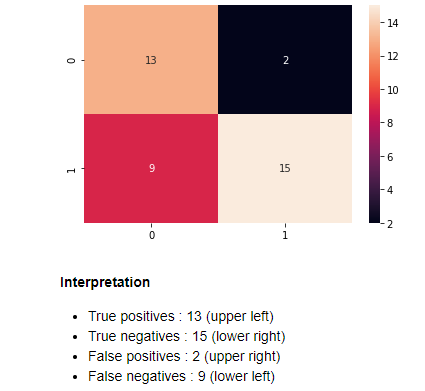
**From model evaluation these are the findings which we have noticed**

* From the model developed, R2 value demonstrates that in 99.7 percent of the cases stepcount is explained by the input exercise attributes (caloriesburnt, sleephrs , weight, mood\_Happy, mood\_Neutral, mood\_Sad, activeness\_Active, activeness\_In-active)
* It is observed that the value of root mean squared error is 114.4, which is lying within the limit of min and max for stepcount while running describe() function in Python Programming language.
* This means algorithm is accurate but can make good predictions

**Our second objective was to find  Does exercise/ working-out improve a person’s activeness?**

Here we have done naïve bayesclassification algorithm between persons activeness as the dependent variable and calories burned, sleep hours and mood as independent variables. Here we are not considering all variables in the data set because we want to satisfy the condition for naïve bayes that all input features should be independent .from the scatter plot between step count and calories burned we have observed that there is a strong correlation between step count and calories burned so, we dropped step count column. Weight also we are not considering here because our data is based on a same person.

For evaluating our model we have tested accuracy score and confusion matrix .From the naïve bayesalgorithm we got an accuracy of 71.79%. below is our confusion matrix



**Usage:**

* We can find the association between exercise and body fitness
* In health care sector we can use this model to predict a person’s fitness