IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115105

Calories Burnt Prediction Using Machine Learning

Rachit Kumar Singh[1], Vaibhav Gupta[2]

[1,2]Student, Information Technology, Maharaja Agrasen Institute of Technology, Rohini, Delhi

Abstract: Machine Learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build models and employ algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. These models can be applied in different areas and trained to match the expectations of management so that accurate steps can be taken to achieve the organization's target.The object of this research paper is to create a project that can be used predict calories burnt using Machine Learning with Python. Xgboost Regression model is used in this project.

INTRODUCTION

The carbohydrates are broken into glucose and converted into energy using oxygen. The muscles that are doing exercise need more oxygen and as the body requires more oxygen the heart beat will increase a lot, so increased heart beat means increased blood flow which in turn will give more oxygen to the muscle which is used to break those glucose molecules , so the energy from these glucose molecules is used and when it is used only a part is used and rest is converted into heat , so body temperature would increase and our body will sweat, so the parameters which we would be taking into consideration for input are :- Duration for which the exercise is done, Average heart beat per minute, Body temperature, Height , weight and gender of the person .
All these would be used to create a prediction model also for calories burnt.

TECHNOLOGY USED

Xgboost Regressor:
Xgboost regressor has two parameters lambda(λ) which denotes Regulation parameter , more the regulator parameter more is the tuning of the decision tree and the other parameter is gamma(γ) which is threshold.

If Similarity Weight(SM) =Σ(Residue)^2/(No of Residues + λ)
Than Gain = Similarity Weight(left decision tree) + Similarity Weight(right decision tree) –

Similarity Weight(root).
If(Gain > γ) than decision tree bifurcation takes place for further levels else not takes place, this makes the xgboost algorithm efficient as compared to others.

Following are the features of support vector machine

- It is the most famous algorithm of xgboost.
- Tianqi-Chan was the founder of xgboost.
- It is platform free.
- It is integrable with multiple systems.
- Xgboost has high speed of processing.
- Xgboost uses parallelization, uses maximum available computational power of the system.
- Xgboost keeps all intermediate calculations in cache so that we don't have to do the same calculation again and again.
- If we have data such that size of our data is more than the size of the memory than xgboost optimized data can work on data greater than the size of the RAM.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115105

Linear Regression

Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y).

More specifically, that y can be calculated from a linear combination of the input variables (x). When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.

A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0).

Logistic Regression

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable.

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.

Logistic regression, despite its name, is a classification model rather than regression model.

$$\log\left(\frac{p}{1-p}\right) = \beta_o + \beta(\text{Age})$$

Lasso Regression

Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of muticollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

Lasso solutions are quadratic programming problems, which are best solved with software (like Matlab). The goal of the algorithm is to minimize:

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Which is the same as minimizing the sum of squares with constraint Σ |Bj≤ s (Σ= summation notation). Some of the βs are shrunk to exactly zero, resulting in a regression model that's easier to interpret.

Gradient Descent

- Gradient Descent is an optimization algorithm used for minimizing the loss function in various machine learning algorithms. It is used for updating the parameters of the learning model.
- w = w - α*dw
- b = b - α*db

METHODOLOGY

- The project starts with importing numpy, pandas, matplotlib, pylot, sns ,xgboost regressor and other libraries.
- The data is analysed that the heart rate and body temperature would be more when the person is doing exercise.
- The data is than visualized using distribution graphs and sns library is used to give grid lines during plot. We will find the distribution of density v/s age, distribution of density v/s height and many more.
- Than we will find the correlation in the dataset by constructing heat map.
- Than we will train the model using xgboost regressor .
- Evaluate the model based on test data and compare the data with original value and find the mean absolute error.
- Than we will make a prediction model on training data , built a predictive system based on which we can predict output from a single input.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115105

Steps For Linear Regression:-

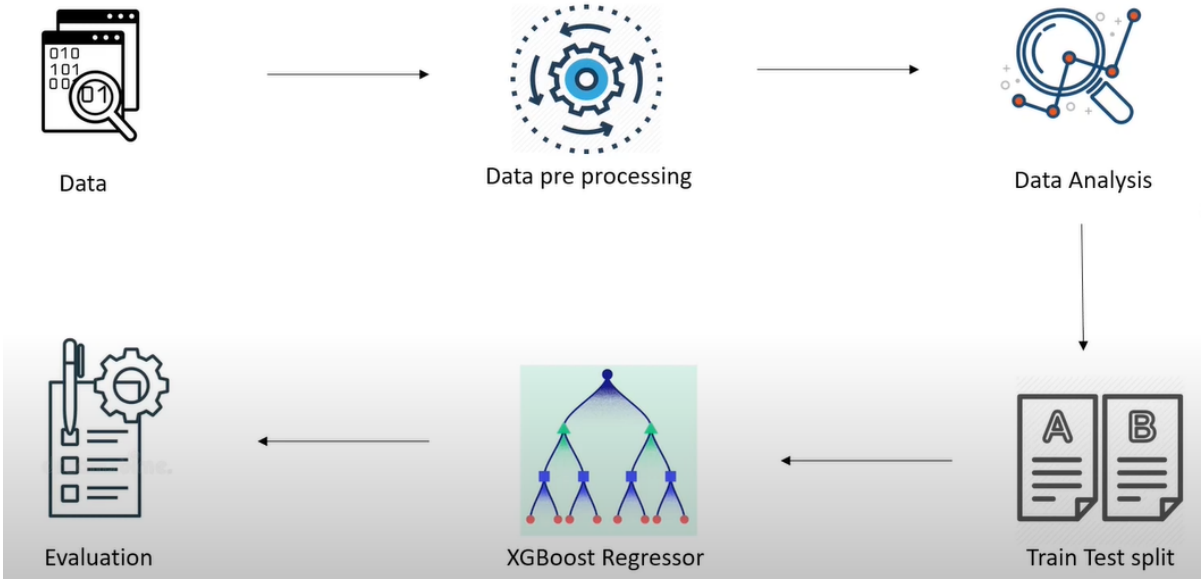
- Than we will train the model using Linear Regressor .
- Evaluate the model based on test data and compare the data with original value and find the mean absolute error and R square value.
- Than we will make a prediction model on training data , built a predictive system based on which we can predict output from a single input.

Steps For Logistic Regression:-

- Than we will train the model using Logistic Regressor .
- Evaluate the model based on test data and compare the data with original value and find the mean absolute error and R square value.
- Than we will make a prediction model on training data , built a predictive system based on which we can predict output from a single input.

Steps For Lasso Regression:-

- Load the required modules and libraries
- Load and analyse the dataset given in the problem statement
- Create training and test dataset
- Build the model and find predictions for the test dataset
- Evaluate the lasso model



RESULTS

Finding distribution of height column
Data Visualization :- find distribution of gender column for example how many males in this data point and how many females in this data point in count-plot
training the model with X_train, my xgboost regressor will learn from the model automatically
prediction on training data

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

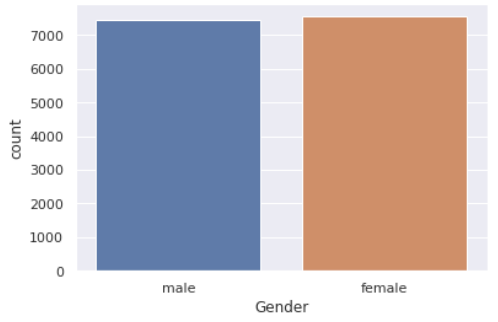
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115105

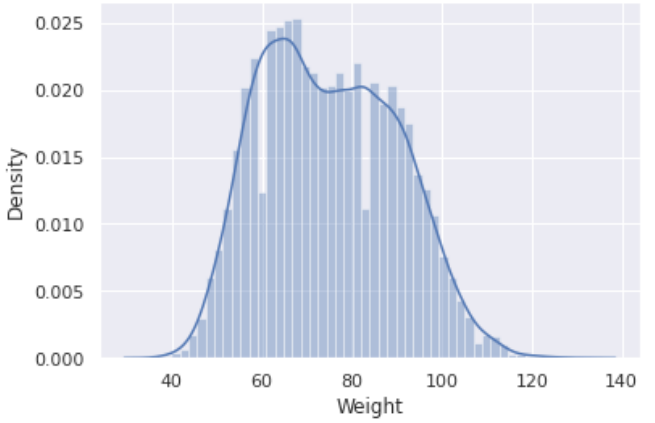
```
sns.countplot(calories_data['Gender'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_deco
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fdf6c22
```




```
#finding distribution of height column
sns.distplot(calories_data['Weight'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distri
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fdf5fb42
```



compare the values predicted by our model with the original values, the original values are the ytest building the predictive system based on input training my model with lasso regression

```
model.fit(X_train, Y_train)
XGBRegressor(copy_X=True, fit_intercept=True,
             n_jobs=None, normalize=False)
```

```
mae = metrics.mean_absolute_error
(Y_test,test_data_prediction)
```

```
print("Mean Absolute Error=", mae)
```

```
Mean Absolute Error= 2.7159012502233186
```

```
input_data = (1,20,166.0,60.0,14.0,94.0,40.3)
#changing tuple data type above to numpy array
input_data_as_numpy_array = np.asarray(input_data)
input_data_reshaped = input_data_as_numpy_array.
reshape(1,-1)
prediction = model.predict(input_data_reshaped)
print(prediction)
print('The predicted calorie burnt is :-', prediction[0])
```

```
[64.3578]
The predicted calorie burnt is :- 64.3578
```

```
X_train1, X_test1, Y_train1, Y_test1 =
train_test_split(X.values,Y.values,
test_size=0.33, random_state = 2)
model1 = Lasso()
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115105

Model fitting for lasso regression

```
model1.fit(X_train1, Y_train1)

Lasso()

test_data_prediction10 = model1.predict(X_test1)
```

finding the mean absolute error for lasso regression.

```
print(test_data_prediction10)
mae = metrics.mean_absolute_error
(Y_test1,test_data_prediction10)
print(mae)
```

finding the mean absolute error for logistic regression.

```
X_train2, X_test2, Y_train2, Y_test2 =
train_test_split(X.values,Y.values,
test_size=0.33, random_state = 2)
print(X.values.shape, X_train2.shape, X_test2.shape)

model2 = LogisticRegression()
```

Training model with linear regression, fitting the model and finding mean absolute error.

```
X_train3, X_test3, Y_train3, Y_test3 =
train_test_split(X.values,Y.values,
test_size=0.33, random_state = 2)

model3.fit(X_train3, Y_train3)

LinearRegression()

X_test_prediction11 = model3.predict(X_test3)

print(X_test_prediction11)

[137.35088567 182.22802451  50.21587504 ...  2
 122.66575284]

mae = metrics.mean_absolute_error(Y_test3,X_te
print("Mean error is", mae)
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115105

CONCLUSION

In this paper, we described xgboost algorithm, linear regression, logistic regression and lasso regression and how they can be used to implement a algorithm for finding the concrete calories burnt which depend on a number of factors. We trained our model than found our output on test data and ultimately found the mean absolute error so as to see the accuracy of our model. Finally we created a real time predictive model, which could be used to find output from given inputs. As per our conclusion xgboost regressor is the best model than can be used in this prediction. to machine learning. Cambridge

REFERENCES

[1] Smola, A., & Vishwanathan, S. V. N. (2008).
[2] Saltz, J. S., & Stanton, J. M. (2017). An introduction to data science Sage Publications.
[3] Shashua, A. (2009). Introduction to machine learning: Class notes 67577. arXiv preprint arXiv:0904.3664.
[4] MacKay, D. J., & Mac Kay, D. J. (2003).algorithms. Cambridge university press.
[5] Daumé III, H. (2012). A course in machine ear learning. Publisher, ciml. info, 5, 69.Introduction
[6] Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.
[7] Cerrada, M., & Aguilar, J. (2008).Reinforcement
[8] Welling, M. (2011). A first encounter with Machine Learning. Irvine, CA.: University of California, 12.
[9] Learning, M. (1994). Neural and Statistical Classification. Editors D. Mitchie et. al, 350.
[10] Mitchell, T. M. (1999). Machine learning and data mining. Communications of the ACM,42(11), 30-36.
[11] Downey, A. B. (2011). Think stats. "O'Reilly Media, Inc.".

