Article

Indication of Health Status – BMI (based on machine learning)

Abdul Hannan 1, Syed Shahroz Abbas Shah 2 and Mazhar Javaid Awan 3

|  |
| --- |
| **Citation:** Lastname, F.; Lastname, F.; Lastname, F. Title. *Appl. Sci.* **2021**, *11*, x. https://doi.org/10.3390/xxxxx  Academic Editor: Firstname Lastname  Received: date  Accepted: date  Published: date  **Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.    **Copyright:** © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). |

1 Department of Software Engineering, University of Management and Technology, Lahore 54770, Pakistan; abdulhanan7877@gmail.com

2 Department of Software Engineering, University of Management and Technology, Lahore 54770, Pakistan; shahroz.abbas99@gmail.com

3 Department of Computer Sciences, University of Management and Technology, Lahore 54770, Pakistan; mazhar.awan@umt.edu.pk

**Abstract:** In today’s modern world obesity has become a major problem and everyone wants to get fit. Hence, to decide on a proper diet plan and health status we need BMI (Body Mass Index) to enquire the level of obesity. We will tell the user his/her health status by comparing the values entered by the user with our data set. Our data set consists of adults less than or equal to seventy years and will give about 98% accurate results. Our model uses BMI to calculate the health status i.e., Weak, Normal, Overweight and etc. The BMI is calculated using the age, gender, height, and weight of the user. The proposed study uses a machine learning approach (models) to evaluate the BMI.

**Keywords:** BMI; Prediction of Health; Health Status by BMI; Obesity Level

1. Introduction

Obesity has become one of the major concern or problem of every third individual out of 10. It can lead towards various diseases and risks such as Diabetes, Heart problems, Asthma and many others such type of diseases. Obesity has different levels of which are measured by its indexes i.e., 0, 1, 2, 3, 4, 5 (Extremely Weak, Weak, Normal, Overweight, Obesity and Extreme obesity). To measure obesity level BMI (Body Mass Index) is considered most accurate and most reliable method [1]. BMI can be calculated by using weight and height of an individual.

BMI, is defined as body weight measured in kilograms divided by square of height in meters [2]:

mass(kg)

BMI = \_\_\_\_\_\_\_\_

height(m) 2

**Indexes and their related terms:**

* 0 – Extremely Weak
* 1 – Weak
* 2 – Normal
* 3 – Overweight
* 4 – Obesity
* 5 – Extreme Obesity

And also, we have:

Gender: Male/Female

Height: Number (cm)

Weight: Number (kg)

One of the main causes of obesity is the high intake of calories and less burn of calories due to lack of exercise or workout. BMI can be use as an indicator which tells us that is our body weight healthy or unhealthy and if it is unhealthy then corresponding diseases for instance heart diseases, diabetes and other relevant diseases.

For statistical approach we are using Model selection and Multiple linear regression model.

2. Materials and Methods

We will be using the following algorithms/methods/models (machine learning) for evaluating and calculating the BMI.

2.1 Models:

* Linear Regression
* Random Forest

2.2 Dataset:

We have taken a random dataset of 500 individuals including men and women by collecting their respective heights and weights, genders and indexes from Kaggle [3].

Figure – 1.1 below shows the header of our dataset

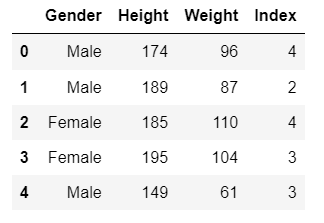


Figure – 1.1

Figure – 1.2 below shows mean, standard deviation, and various percentiles.

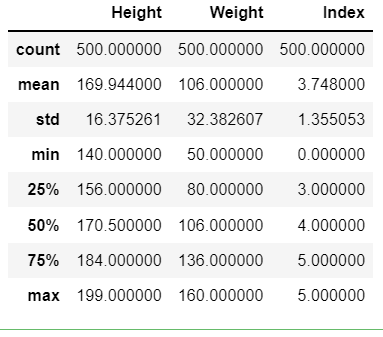


Figure – 1.2

Figure – 1.3 below shows correlation of our dataset entities.

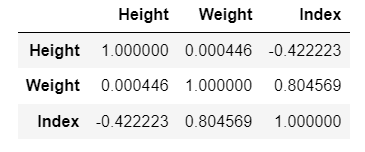


Figure – 1.3

Figure – 1.4 below shows that there is no missing info (value) in our dataset.

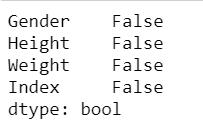


Figure – 1.4

Figure – 1.4.1 below shows that there is no missing info (value) in our dataset by using heatmap.

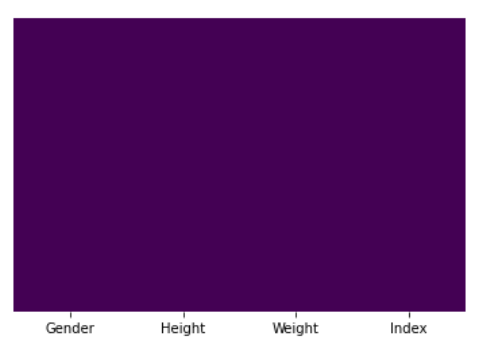


Figure – 1.4.1

Figure – 1.5 below shows that what the health status of persons in our dataset.

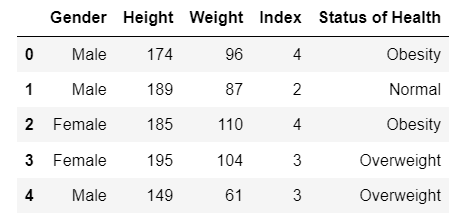


Figure – 1.5

Figure – 1.6 below represents the gender with label (0, 1).

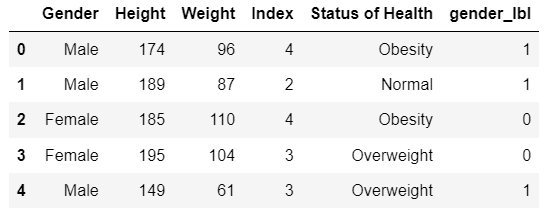


Figure – 1.6

Figure – 1.7 below shows that how many males and females are present in our dataset.

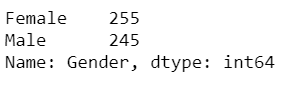


Figure – 1.7

Figure – 1.8 below shows that how many persons have which type of health status.

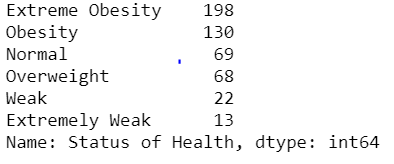


Figure – 1.8

Figure – 1.9 below shows the categories of gender with pie plot.

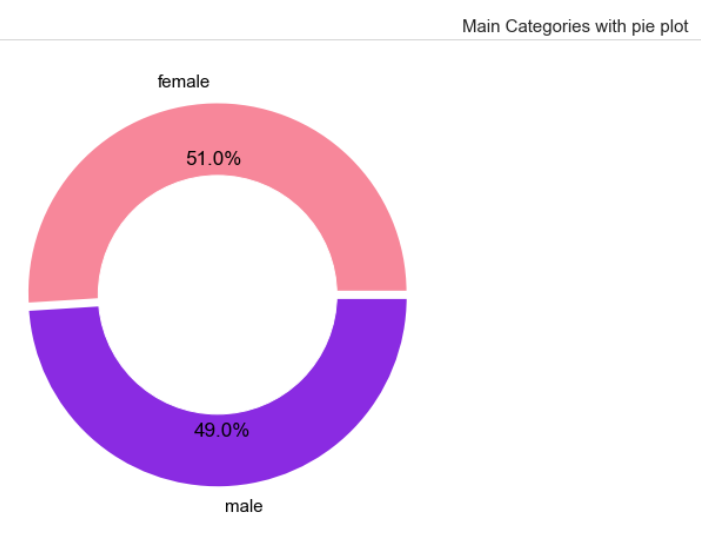


Figure – 1.9

Figure – 1.9.1 below shows the status of health with pie chart.

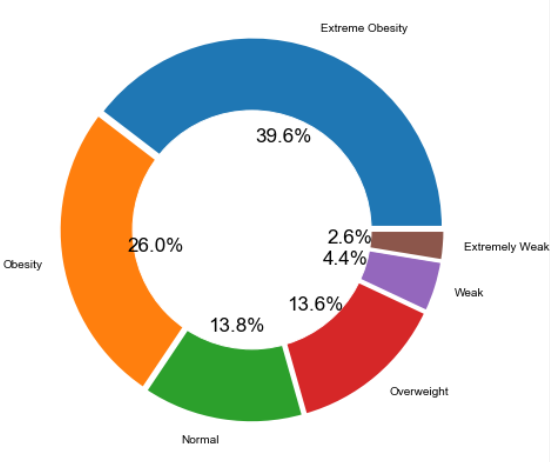


Figure – 1.9.1

Figure – 1.10 below compare the Male and Female health status.

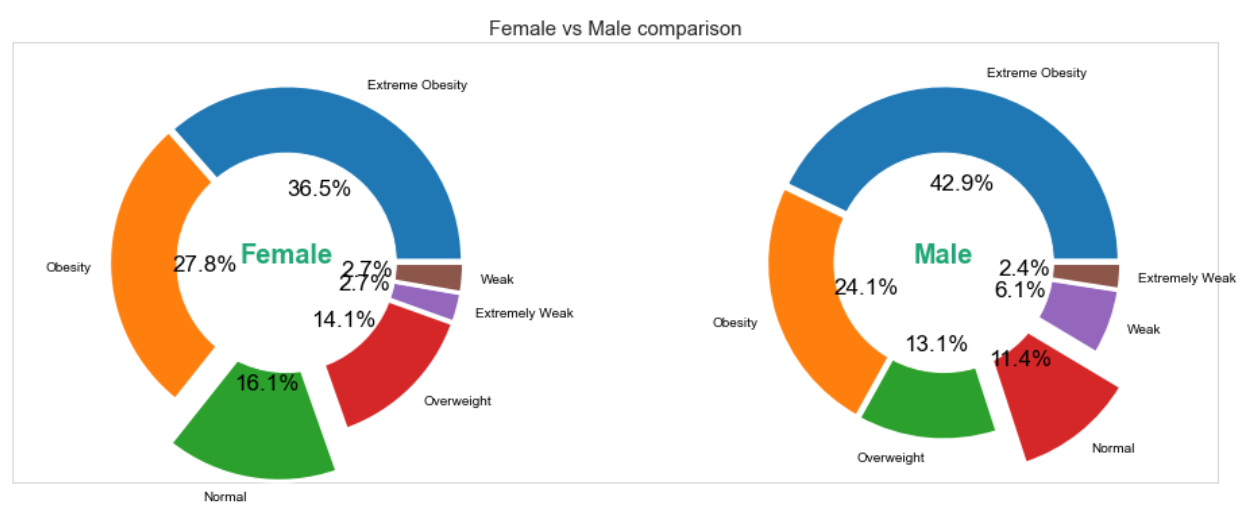


Figure – 1.10

Figure – 1.11 shows the decision tree.

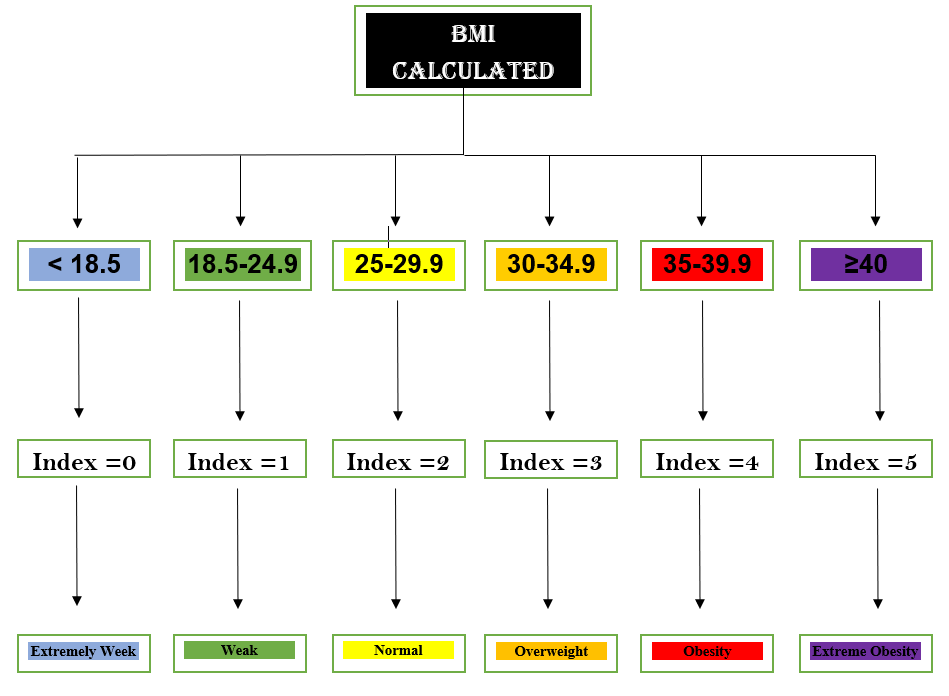


Figure – 1.11

3. Implementation:

After receiving the data from our data set, our algorithm will read the corresponding data values (gender, height, weight) and perform the specified calculations on given entities and evaluate BMI.

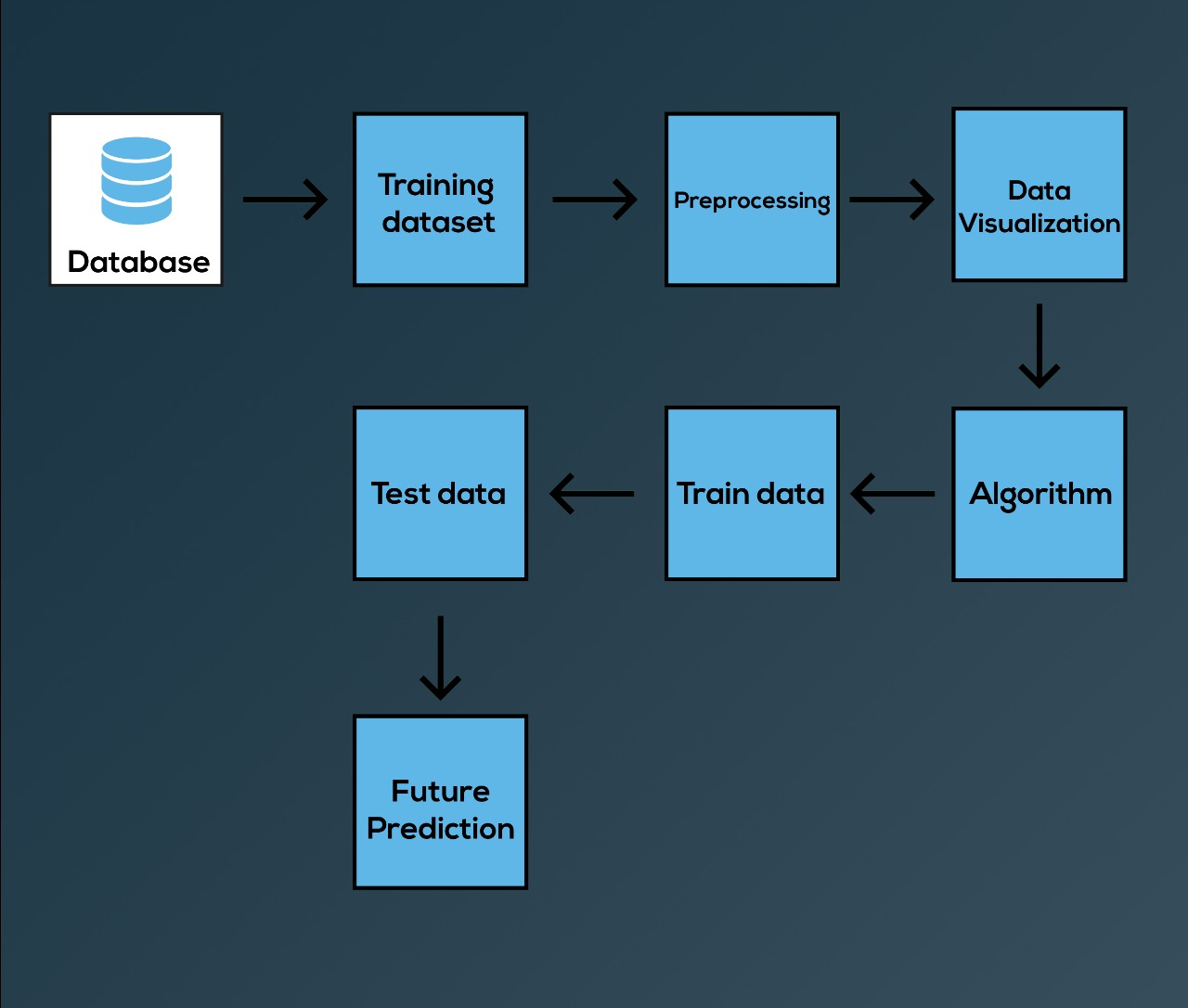
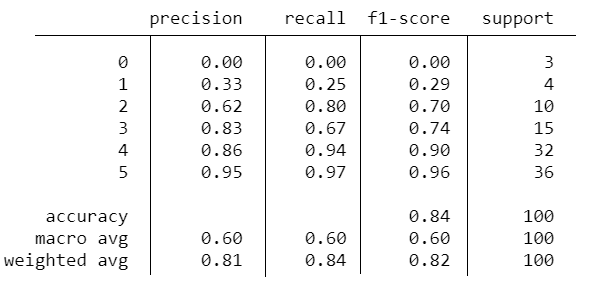
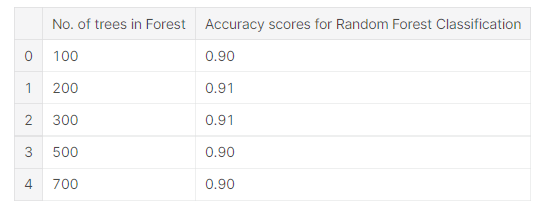


Figure - 2 [2-vis]

3. Results



**Table 1.** This table shows precision and accuracy of Linear Regression model which we applied on our data set.



**Table 2.** This table shows precision and accuracy of Random Forest model which we applied on our data set.

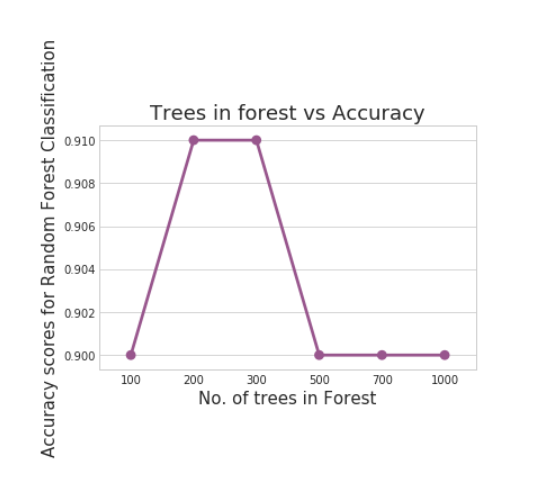


Figure – 3.1

|  |  |  |
| --- | --- | --- |
| **Models** | **Data** | **Accuracy(%)** |
| Linear Regression | 500 persons data | 84 |
| Random Forest | 500 persons data | 91% |
|  |  |  |

4. Discussion

Obesity being a very common problem in today’s modern world lead towards many serious diseases i.e., Diabetes, Heart problems, Asthma and others. So, in order to know the health status and level of obesity BMI is used which consist of a person’s height and weight.

In this study, we have experimented with BMI calculation using a data set of 500 individual (without missing any value) using Random Forest model and Linear Regression model. The dataset, which we used, was collected, labeled, and trained by devbrabal who is a student in at National Institute of Technology, Raipur Raipur, Chhattisgarh, India and we have taken this data set from his kaggel account.

We compared the performances of Liner Regression and Random Forest and achieved a good result. However, Random Forest performed better than Linear Regression by giving 91% accuracy rate whereas Linear Regression gave an accuracy of 84%.  
Moreover, the aim of our study was to provide an accurate result by calculating the BMI of the person after he/she enters his/her data We collected data for adult males and females and transgender were omitted from this study.  
In future work we can further improve the accuracy by adding more samples and increasing the range of BMI such as for children and people in old ages as well. Furthermore, the study could also include transgender in it since this category is also widely spread and can also be benefitted from this model in future.

**Author Contributions:** “Conceptualization, A.H. and S.A.; methodology, S.A.; software, A.H.; validation, A.H. and S.A.; formal analysis, A.H.; investigation A.H.; resources, A.H.; data curation, A.H. and S.A.; writing—original draft preparation, S.A.; writing—review and editing, S.A.; visualization, A.H.; supervision, S.A.; project administration, S.A.; funding acquisition, None. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding

**Institutional Review Board Statement:** Not Applicable for this study.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data set used is available at following google drive link:

<https://drive.google.com/drive/folders/1DlL_Jttby-P7zZrVojRTiWxt7gSK-Q6f?usp=sharing>

**Conflicts of Interest:** The authors declare no conflict of interest.

References

1. <https://www.researchgate.net/publication/269417406_Statistical_methods_for_body_mass_index_a_selective_review_of_the_literature>
2. <https://www.researchgate.net/publication/269417406_Statistical_methods_for_body_mass_index_a_selective_review_of_the_literature>
3. <https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex?select=500_Person_Gender_Height_Weight_Index.csv>
4. 2-vis <https://www.google.com/search?q=BMI+calculation+pics+datascience+machine+learning&tbm=isch&ved=2ahUKEwi2i6fE9eHxAhULShoKHfXdAcsQ2-cCegQIABAA&oq=BMI+calculation+pics+datascience+machine+learning&gs_lcp=CgNpbWcQA1CjCliJc2D4dWgAcAB4AIABjwKIAfw1kgEEMi0yOZgBAKABAaoBC2d3cy13aXotaW1nwAEB&sclient=img&ei=F4LuYPajGYuUafW7h9gM&bih=577&biw=1280#imgrc=TLxjOUDofGUSCM&imgdii=AoLhVKnVaAQjBM>
5. 1-vis <https://www.google.com/search?q=BMI+calculation+pics+datascience+machine+learning&tbm=isch&ved=2ahUKEwi2i6fE9eHxAhULShoKHfXdAcsQ2-cCegQIABAA&oq=BMI+calculation+pics+datascience+machine+learning&gs_lcp=CgNpbWcQA1CjCliJc2D4dWgAcAB4AIABjwKIAfw1kgEEMi0yOZgBAKABAaoBC2d3cy13aXotaW1nwAEB&sclient=img&ei=F4LuYPajGYuUafW7h9gM&bih=577&biw=1280#imgrc=TLxjOUDofGUSCM>