**Title: Prediction of Obesity Levels Using Data Mining Techniques**

Chalumuri, Tejalakshmi Gangadhar

MS in Data Science, Seidenberg School of CSIS

CS619: Data Mining

Professor**:** D. Paul Benjamin

Submission Date: May 9, 2024

**Introduction:**

The research project is about developing predictive models to examine obesity levels by using data mining techniques on dietary habits, body conditions and other factors. I have applied several classification algorithms to predict the levels of obesity in people from three countries: Mexico, Peru, and Colombia, by employing a dataset that entails 2,111 examples with 17 different attributes. These algorithms were supposed to provide a better understanding of the impact of various lifestyle factors on obesity for this population to take care of the vulnerable sections.

**Dataset Description:**

The analyzed dataset contains 2,111 entries each having numerous columns relating to subjects’ eating behavior, health condition, and personal particulars. These are then further subdivided into specific stages of being overweight thereby giving a clear picture about the matter at hand. The information in the dataset was compiled through synthetic methodology combined with real data collection via a web platform. This methodological approach ensures that we have strong enough database which can be used in our predictive models aimed at finding out all variables affecting or contributing towards high obesity rates accurately and predicting their impacts on future cases.

**Data Preparation:**

I imported the dataset used in this study, using the explorer feature of WEKA tool. There were several preprocessing steps that were undertaken to ensure that the data was well prepared for analysis. Through normalization process, numerical attributes in the dataset were scaled, thus ensuring a consistent scale and no bias towards results. Additionally, encoding techniques were used for categorical variables. That was needed to convert these variables into a format that could be easily processed by classification algorithms ensuring more accurate and efficient analysis. These actions are essential for maintaining data integrity and credibility while doing predictive modeling.

**Classification Algorithms Used:**

* Random Forest
* Logistic Regression
* J48- Decision Tree

**Data Analysis and Results:**

I’ve employed three different classification methods to predict obesity levels and measured their effectiveness with various performance indicators such as accuracy, precision, recall, F-measure, The Matthews Correlation Coefficient (MCC), and Area under receiver operating characteristic (ROC) curve. In order to enable us deeply to analyze them all and look how well they worked compared with one another we got these evaluation results from Weka software package. The scope of this assessment will help us appreciate both strong points of those approaches as well as their limitations when it comes to accurate forecasting of obesity.

**Random Forest:**

The output is obtained from the Random Forest classification using Weka. Out of 422 instances correctly classified instances are 402 and incorrectly classified instances are 20. Below is the summary of the model:

* **Accuracy**: 95.2607%
* Correctly Classified Instances**:** 402(95.2607 %)
* Incorrectly Classified Instances: 20(4.7393 %)
* Kappa statistic: 0.9444
* Mean absolute error: 0.0481.
* Root mean squared error: 0.1222
* Relative absolute error: 19.6642 %
* Root relative squared error: 34.9431 %
* Total Number of Instances: 422
* Precision, Recall, F-Measure, MCC, and ROC Area for key classes:
  + **Normal Weight:** Precision 0.880, Recall 0.957, F-Measure 0.917, MCC 0.901, ROC Area 0.994
  + **Obesity Type I:** Precision 1.000, Recall 0.986, F-Measure 0.993, MCC 0.992, ROC Area 1.000

**A screenshot of a computer

Description automatically generated**

**Fig1: Summary (Random Forest)**

(Source: Acquired From WEKA)

**A screenshot of a computer

Description automatically generated**

**Fig2: Detailed Accuracy by Class (Random Forest)**

(Source: Acquired From WEKA)

**A screenshot of a computer code

Description automatically generated**

**Fig3: Confusion Matrix (Random Forest)**

(Source: Acquired From WEKA)

**Logistic Regression:**

The output is obtained from the Logistic Regression classification using Weka. Out of 422 instances correctly classified instances are 405 and incorrectly classified instances are 17. Below is the summary of the model:

* **Accuracy**: 95.9716%
* Correctly Classified Instances; 405(95.9716 %)
* Incorrectly Classified Instances: 17(4.0284 %)
* Kappa statistic:0.9528
* Mean absolute error: 0.0129
* Root mean squared error: 0.1061
* Relative absolute error :5.2828 %
* Root relative squared error: 30.3413 %
* Total Number of Instances: 422
* Precision, Recall, F-Measure, MCC, and ROC Area for key classes:
  + Normal Weight: Precision 0.955, Recall 0.913, F-Measure 0.933, MCC 0.921, ROC Area 0.995
  + Obesity Type I: Precision 1.000, Recall 0.958, F-Measure 0.978, MCC 0.974, ROC Area 0.989

A screenshot of a computer

Description automatically generated

**Fig4: Summary (Logistic Regression)**

(Source: Acquired From WEKA)

A screenshot of a computer screen

Description automatically generated

**Fig5: Detailed Accuracy by Class (Logistic Regression)**

(Source: Acquired From WEKA)

A screenshot of a computer code

Description automatically generated

**Fig6: Confusion Matrix (Logistic Regression)**

(Source: Acquired From WEKA)

**J48 Decision Tree:**

The output is obtained from the J48 Decision Tree classification using Weka. Out of 422 instances correctly classified instances are 393 and incorrectly classified instances are 29. Below is the summary of the model:

* **Accuracy**: 93.128%
* Correctly Classified Instances: 393 (93.128 %)
* Incorrectly Classified Instances: 29( 6.872 %)
* Kappa statistic: 0.9195
* Mean absolute error: 0.0234
* Root mean squared error: 0.136
* Relative absolute error: 9.5575 %
* Root relative squared error: 38.8751 %
* Total Number of Instances: 422
* Precision, Recall, F-Measure, MCC, and ROC Area for key classes:
  + Normal Weight: Precision 0.864, Recall 0.826, F-Measure 0.844, MCC 0.815, ROC Area 0.958
  + Obesity Type I: Precision 0.959, Recall 0.986, F-Measure 0.972, MCC 0.967, ROC Area 0.990

A screenshot of a computer

Description automatically generated

**Fig7: Summary (J48 Decision Tree)**

(Source: Acquired From WEKA)

A screenshot of a computer screen

Description automatically generated

**Fig8: Detailed Accuracy by Class (J48 Decision Tree)**

(Source: Acquired From WEKA)

A screenshot of a computer code

Description automatically generated

**Fig9: Confusion Matrix (J48 Decision Tree)**

(Source: Acquired From WEKA)

**Discussion:**

The models I have tested across multiple metrics showed a solid performance, with Logistic Regression standing out for its accuracy and robustness in other evaluations. These models can be useful tools for devising public health strategies by revealing determinants of obesity. In addition I have also gained insights into the suitability of individual models for specific classification tasks by examining their performance variation and different metrics. Therefore, because of these models’ usefulness, they are applied precisely to address and mitigate obesity through public health interventions.

**Conclusions:**

This project demonstrated the strength of data mining techniques in forecasting levels of obesity using several health and lifestyle attributes. Out of all the evaluated models, Logistic Regression was found to be most effective, which suggests that it could be adopted in other similar predictive analytics healthcare projects. Henceforth, this implies that Logistic Regression works so well on such cases; hence would be useful in further investigations as well as future applications seeking to comprehend and predict health outcomes on different sets of data. This strongly indicates its potential adoption across medical analytics where accuracy is vital in making predictions.

**Dataset Source:** [**Kaggle.com**](https://www.kaggle.com/)

**Dataset URL:** [**Obesity Levels**](https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels)

**Appendices:**

**Appendix 1: Random Forest**

=== Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Relation: ObesityDataSet\_raw\_and\_data\_sinthetic-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0

Instances: 2111

Attributes: 17

Age

Gender

Height

Weight

CALC

FAVC

FCVC

NCP

SCC

SMOKE

CH2O

family\_history\_with\_overweight

FAF

TUE

CAEC

MTRANS

NObeyesdad

Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.65 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.04 seconds

=== Summary ===

Correctly Classified Instances 402 95.2607 %

Incorrectly Classified Instances 20 4.7393 %

Kappa statistic 0.9444

Mean absolute error 0.0481

Root mean squared error 0.1222

Relative absolute error 19.6642 %

Root relative squared error 34.9431 %

Total Number of Instances 422

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.957 0.025 0.880 0.957 0.917 0.901 0.994 0.960 Normal\_Weight

0.842 0.014 0.906 0.842 0.873 0.854 0.992 0.965 Overweight\_Level\_I

0.926 0.003 0.980 0.926 0.952 0.946 0.997 0.983 Overweight\_Level\_II

0.986 0.000 1.000 0.986 0.993 0.992 1.000 1.000 Obesity\_Type\_I

0.976 0.005 0.952 0.976 0.964 0.960 0.998 0.985 Insufficient\_Weight

0.984 0.006 0.969 0.984 0.976 0.972 1.000 0.999 Obesity\_Type\_II

0.985 0.003 0.985 0.985 0.985 0.982 1.000 1.000 Obesity\_Type\_III

Weighted Avg. 0.953 0.008 0.953 0.953 0.952 0.945 0.997 0.985

=== Confusion Matrix ===

a b c d e f g <-- classified as

66 1 0 0 2 0 0 | a = Normal\_Weight

8 48 1 0 0 0 0 | b = Overweight\_Level\_I

0 4 50 0 0 0 0 | c = Overweight\_Level\_II

0 0 0 70 0 1 0 | d = Obesity\_Type\_I

1 0 0 0 40 0 0 | e = Insufficient\_Weight

0 0 0 0 0 62 1 | f = Obesity\_Type\_II

1. 0 0 0 0 1 66 | g = Obesity\_Type\_III

**Appendix 2: Logistic Regression**

=== Run information ===

Scheme: weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

Relation: ObesityDataSet\_raw\_and\_data\_sinthetic-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0

Instances: 2111

Attributes: 17

Age

Gender

Height

Weight

CALC

FAVC

FCVC

NCP

SCC

SMOKE

CH2O

family\_history\_with\_overweight

FAF

TUE

CAEC

MTRANS

NObeyesdad

Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8

Coefficients...

Class

Variable Normal\_Weight Overweight\_Level\_I Overweight\_Level\_II Obesity\_Type\_I Insufficient\_Weight Obesity\_Type\_II

===================================================================================================================================================================================

Age 370.8851 798.5293 807.1844 840.6355 434.4701 1308.2132

Gender=Male 356.0947 314.2311 315.1092 288.6157 360.8084 462.633

Height 4572.3601 962.7684 879.2158 330.2161 4915.6044 -454.2722

Weight -14910.5564 -4218.2682 -3966.3147 -2368.9563 -16331.2673 784.2891

CALC=no 27.3904 15.4557 14.7088 10.8839 24.597 71.5062

CALC=Sometimes -76.0787 -53.5452 -57.2006 -59.8793 -73.3509 -66.9503

CALC=Frequently 348.7091 333.7119 334.6596 350.4894 372.8776 -5.2636

CALC=Always 43.8836 -4243.2635 -2251.1858 -355.1306 -1633.2948 42.1436

FAVC=yes 46.5298 41.7812 38.9952 38.8487 59.2498 -74.1295

FCVC -229.6986 -376.8618 -377.2186 -378.3846 -220.4232 -432.5067

NCP -307.5504 -414.6423 -415.8503 -416.2789 -310.0464 -480.5196

SCC=yes -176.1657 32.1292 31.6375 -19.7352 -179.3684 -428.5902

SMOKE=yes 205.0366 -114.4636 -112.2513 -98.9835 202.6788 -11.2267

CH2O 48.0679 97.6849 97.7993 92.2569 58.4127 -81.0004

family\_history\_with\_overweight=no -132.4613 -129.1203 -131.9498 -132.407 -134.3174 38.7603

FAF 221.9703 64.2126 62.2745 39.54 238.3227 -113.4394

TUE 79.1649 92.5773 95.1912 98.6137 75.4111 90.4415

CAEC=Sometimes -59.3606 21.7324 22.1255 41.4956 -64.3643 37.9465

CAEC=Frequently -68.3658 -104.0143 -103.7327 -115.1017 -52.9851 -243.4396

CAEC=Always 513.9187 386.043 380.4373 399.1226 481.4472 412.9193

CAEC=no 105.8201 -79.322 -76.9996 -160.047 102.4014 398.7717

MTRANS=Public\_Transportation -58.7183 -49.2103 -34.6532 -30.7895 -39.187 14.906

MTRANS=Walking 90.3071 144.1453 151.3661 148.7036 133.0069 69.2391

MTRANS=Automobile -3.2662 -9.7527 1.7966 -8.3096 22.3728 -52.4719

MTRANS=Motorbike 1342.5149 924.7413 936.2254 943.971 279.8779 854.8714

MTRANS=Bike 701.2298 727.0905 -769.3257 -462.0945 -394.1841 -37.9579

Intercept 2607.3228 1717.7275 1672.8269 1358.1758 2544.7069 184.4043

Odds Ratios...

Class

Variable Normal\_Weight Overweight\_Level\_I Overweight\_Level\_II Obesity\_Type\_I Insufficient\_Weight Obesity\_Type\_II

===================================================================================================================================================================================

Age 1.1839927163589935E161 Infinity Infinity Infinity 4.87507348988398E188 Infinity

Gender=Male 4.46654525021917E154 2.943187256562813E136 7.082281760175523E136 2.2090623745507302E125 4.978535413036093E156 8.298107705733374E200

Height Infinity Infinity Infinity 2.5764651838402003E143 Infinity 0

Weight 0 0 0 0 0 Infinity

CALC=no 7.86141502927269E11 5156335.0672 2443096.8408 53308.9672 4.812158193374473E10 1.1342994506973572E31

CALC=Sometimes 0 0 0 0 0 0

CALC=Frequently 2.7696958337844065E151 8.49679010525376E144 2.1919394873450177E145 1.6428231180914864E152 8.683578516744978E161 0.0052

CALC=Always 1.1439918937824508E19 0 0 0 0 2.00791044468488448E18

FAVC=yes 1.6130071478433605E20 1.39749964891518541E18 8.6177409542005728E16 7.4435100137984496E16 5.393458584271927E25 0

FCVC 0 0 0 0 0 0

NCP 0 0 0 0 0 0

SCC=yes 0 8.985126486596523E13 5.495315154352433E13 0 0 0

SMOKE=yes 1.112448783324171E89 0 0 0 1.052693388661881E88 0

CH2O 7.51000938625114E20 2.6546869128422828E42 2.9763945847122276E42 1.1658907234730537E40 2.3352350398528764E25 0

family\_history\_with\_overweight=no 0 0 0 0 0 6.8138761100954576E16

FAF 2.514529635849348E96 7.712155011075642E27 1.1103429308203995E27 1.48589041788224512E17 3.1786174873680395E103 0

TUE 2.4036520929069355E34 1.606228420345244E40 2.1928100449382337E41 6.720231474052202E42 5.6316640282416406E32 1.8978333721147995E39

CAEC=Sometimes 0 2743290988.3456 4064342362.7342 1.05025454289890675E18 0 3.0195061738920956E16

CAEC=Frequently 0 0 0 0 0 0

CAEC=Always 1.5561576189465795E223 4.532399381091149E167 1.666585961876747E165 2.171361039086144E173 1.2298300936242975E209 2.130914505284053E179

CAEC=no 9.059313062532933E45 0 0 0 2.9672377218198155E44 1.528863600198472E173

MTRANS=Public\_Transportation 0 0 0 0 0 2975745.0263

MTRANS=Walking 1.659079969274987E39 3.9948897637265884E62 5.463131184723859E65 3.8119375764201515E64 5.809810249779235E57 1.1753516159629497E30

MTRANS=Automobile 0.0382 0.0001 6.0289 0.0002 5204322176.3146 0

MTRANS=Motorbike Infinity Infinity Infinity Infinity 3.543355495554408E121 Infinity

MTRANS=Bike 3.4692798552844495E304 Infinity 0 0 0 0

Time taken to build model: 39.2 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances 405 95.9716 %

Incorrectly Classified Instances 17 4.0284 %

Kappa statistic 0.9528

Mean absolute error 0.0129

Root mean squared error 0.1061

Relative absolute error 5.2828 %

Root relative squared error 30.3413 %

Total Number of Instances 422

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.913 0.008 0.955 0.913 0.933 0.921 0.995 0.960 Normal\_Weight

0.930 0.011 0.930 0.930 0.930 0.919 0.985 0.874 Overweight\_Level\_I

0.981 0.005 0.964 0.981 0.972 0.968 0.996 0.927 Overweight\_Level\_II

0.958 0.000 1.000 0.958 0.978 0.974 0.989 0.982 Obesity\_Type\_I

0.976 0.008 0.930 0.976 0.952 0.947 0.996 0.926 Insufficient\_Weight

0.984 0.011 0.939 0.984 0.961 0.955 0.986 0.940 Obesity\_Type\_II

0.985 0.003 0.985 0.985 0.985 0.982 0.988 0.973 Obesity\_Type\_III

Weighted Avg. 0.960 0.006 0.960 0.960 0.960 0.953 0.990 0.944

=== Confusion Matrix ===

a b c d e f g <-- classified as

63 3 0 0 3 0 0 | a = Normal\_Weight

2 53 2 0 0 0 0 | b = Overweight\_Level\_I

0 1 53 0 0 0 0 | c = Overweight\_Level\_II

0 0 0 68 0 3 0 | d = Obesity\_Type\_I

1 0 0 0 40 0 0 | e = Insufficient\_Weight

0 0 0 0 0 62 1 | f = Obesity\_Type\_II

1. 0 0 0 0 1 66 | g = Obesity\_Type\_III

**Appendix 3: J48 Decision Tree**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: ObesityDataSet\_raw\_and\_data\_sinthetic-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0

Instances: 2111

Attributes: 17

Age

Gender

Height

Weight

CALC

FAVC

FCVC

NCP

SCC

SMOKE

CH2O

family\_history\_with\_overweight

FAF

TUE

CAEC

MTRANS

NObeyesdad

Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

------------------

Weight <= 0.451724

| Weight <= 0.156716

| | Height <= 0.396226

| | | Weight <= 0.057132

| | | | Height <= 0.119249: Normal\_Weight (4.0)

| | | | Height > 0.119249

| | | | | Weight <= 0.044166: Insufficient\_Weight (72.0)

| | | | | Weight > 0.044166

| | | | | | Height <= 0.245283: Normal\_Weight (4.0)

| | | | | | Height > 0.245283: Insufficient\_Weight (12.0)

| | | Weight > 0.057132

| | | | Height <= 0.165604

| | | | | FAVC = no: Normal\_Weight (2.0)

| | | | | FAVC = yes

| | | | | | SCC = no

| | | | | | | Weight <= 0.129157: Normal\_Weight (5.0)

| | | | | | | Weight > 0.129157: Overweight\_Level\_I (4.0/1.0)

| | | | | | SCC = yes: Overweight\_Level\_I (6.0)

| | | | Height > 0.165604

| | | | | Weight <= 0.087048

| | | | | | Height <= 0.293215: Normal\_Weight (22.0)

| | | | | | Height > 0.293215: Insufficient\_Weight (8.0/1.0)

| | | | | Weight > 0.087048: Normal\_Weight (84.0/1.0)

| | Height > 0.396226

| | | Weight <= 0.119383: Insufficient\_Weight (124.0)

| | | Weight > 0.119383

| | | | Height <= 0.566762: Normal\_Weight (19.0/1.0)

| | | | Height > 0.566762

| | | | | FAF <= 0.443506: Normal\_Weight (4.0)

| | | | | FAF > 0.443506: Insufficient\_Weight (50.0)

| Weight > 0.156716

| | Weight <= 0.276119

| | | Height <= 0.262453

| | | | FAVC = no: Overweight\_Level\_II (39.0/1.0)

| | | | FAVC = yes

| | | | | Weight <= 0.231343: Overweight\_Level\_I (24.0)

| | | | | Weight > 0.231343: Obesity\_Type\_I (4.0)

| | | Height > 0.262453

| | | | Height <= 0.537419

| | | | | Weight <= 0.246269

| | | | | | Height <= 0.373855

| | | | | | | Weight <= 0.189552: Normal\_Weight (12.0/1.0)

| | | | | | | Weight > 0.189552: Overweight\_Level\_I (85.0/3.0)

| | | | | | Height > 0.373855

| | | | | | | Age <= 0.286686: Normal\_Weight (55.0/2.0)

| | | | | | | Age > 0.286686

| | | | | | | | FCVC <= 0.902767: Overweight\_Level\_I (5.0)

| | | | | | | | FCVC > 0.902767: Normal\_Weight (2.0)

| | | | | Weight > 0.246269

| | | | | | Age <= 0.136989

| | | | | | | Age <= 0.095363: Overweight\_Level\_I (3.0)

| | | | | | | Age > 0.095363: Overweight\_Level\_II (6.0)

| | | | | | Age > 0.136989: Overweight\_Level\_I (56.0)

| | | | Height > 0.537419

| | | | | Height <= 0.754717

| | | | | | Weight <= 0.270946: Normal\_Weight (51.0)

| | | | | | Weight > 0.270946

| | | | | | | Age <= 0.30837: Normal\_Weight (2.0)

| | | | | | | Age > 0.30837: Overweight\_Level\_I (2.0)

| | | | | Height > 0.754717

| | | | | | Weight <= 0.197194: Insufficient\_Weight (6.0)

| | | | | | Weight > 0.197194: Normal\_Weight (4.0)

| | Weight > 0.276119

| | | Height <= 0.526749

| | | | Height <= 0.370547

| | | | | Weight <= 0.395522: Obesity\_Type\_I (122.0/3.0)

| | | | | Weight > 0.395522: Obesity\_Type\_II (4.0)

| | | | Height > 0.370547

| | | | | Weight <= 0.345643

| | | | | | Height <= 0.406036

| | | | | | | Weight <= 0.3163: Overweight\_Level\_II (25.0)

| | | | | | | Weight > 0.3163: Obesity\_Type\_I (11.0)

| | | | | | Height > 0.406036: Overweight\_Level\_II (112.0/1.0)

| | | | | Weight > 0.345643

| | | | | | Weight <= 0.370713

| | | | | | | Height <= 0.462258: Obesity\_Type\_I (15.0)

| | | | | | | Height > 0.462258: Overweight\_Level\_II (7.0)

| | | | | | Weight > 0.370713: Obesity\_Type\_I (64.0/1.0)

| | | Height > 0.526749

| | | | Weight <= 0.38806

| | | | | Height <= 0.65996

| | | | | | Weight <= 0.330322

| | | | | | | Age <= 0.343902: Overweight\_Level\_I (48.0/1.0)

| | | | | | | Age > 0.343902: Overweight\_Level\_II (4.0/1.0)

| | | | | | Weight > 0.330322

| | | | | | | NCP <= 0.684966: Overweight\_Level\_II (54.0)

| | | | | | | NCP > 0.684966: Overweight\_Level\_I (4.0/1.0)

| | | | | Height > 0.65996

| | | | | | Weight <= 0.339537: Normal\_Weight (14.0)

| | | | | | Weight > 0.339537

| | | | | | | CAEC = Sometimes: Overweight\_Level\_I (51.0/1.0)

| | | | | | | CAEC = Frequently: Normal\_Weight (3.0/1.0)

| | | | | | | CAEC = Always: Overweight\_Level\_I (4.0/2.0)

| | | | | | | CAEC = no: Overweight\_Level\_I (0.0)

| | | | Weight > 0.38806

| | | | | Height <= 0.650617: Obesity\_Type\_I (31.0/1.0)

| | | | | Height > 0.650617: Overweight\_Level\_II (33.0/1.0)

Weight > 0.451724

| Gender = Female: Obesity\_Type\_III (324.0/1.0)

| Gender = Male

| | Age <= 0.181362

| | | FCVC <= 0.785637: Obesity\_Type\_I (68.0/3.0)

| | | FCVC > 0.785637

| | | | FAF <= 0.322542: Obesity\_Type\_II (13.0)

| | | | FAF > 0.322542: Obesity\_Type\_I (3.0/1.0)

| | Age > 0.181362

| | | Weight <= 0.526862

| | | | Height <= 0.566762

| | | | | CH2O <= 0.384055: Obesity\_Type\_II (50.0)

| | | | | CH2O > 0.384055: Obesity\_Type\_I (2.0)

| | | | Height > 0.566762

| | | | | Height <= 0.783321: Obesity\_Type\_I (38.0)

| | | | | Height > 0.783321: Overweight\_Level\_II (5.0)

| | | Weight > 0.526862: Obesity\_Type\_II (226.0/1.0)

Number of Leaves : 59

Size of the tree : 115

Time taken to build model: 0.05 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances 393 93.128 %

Incorrectly Classified Instances 29 6.872 %

Kappa statistic 0.9195

Mean absolute error 0.0234

Root mean squared error 0.136

Relative absolute error 9.5575 %

Root relative squared error 38.8751 %

Total Number of Instances 422

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.826 0.025 0.864 0.826 0.844 0.815 0.958 0.836 Normal\_Weight

0.895 0.008 0.944 0.895 0.919 0.907 0.978 0.869 Overweight\_Level\_I

0.944 0.005 0.962 0.944 0.953 0.947 0.979 0.958 Overweight\_Level\_II

0.986 0.009 0.959 0.986 0.972 0.967 0.990 0.965 Obesity\_Type\_I

0.927 0.024 0.809 0.927 0.864 0.850 0.992 0.893 Insufficient\_Weight

0.952 0.006 0.968 0.952 0.960 0.953 0.988 0.959 Obesity\_Type\_II

0.985 0.003 0.985 0.985 0.985 0.982 0.991 0.973 Obesity\_Type\_III

Weighted Avg. 0.931 0.011 0.933 0.931 0.931 0.920 0.982 0.923

=== Confusion Matrix ===

a b c d e f g <-- classified as

57 1 2 0 9 0 0 | a = Normal\_Weight

6 51 0 0 0 0 0 | b = Overweight\_Level\_I

0 2 51 1 0 0 0 | c = Overweight\_Level\_II

0 0 0 70 0 1 0 | d = Obesity\_Type\_I

3 0 0 0 38 0 0 | e = Insufficient\_Weight

0 0 0 2 0 60 1 | f = Obesity\_Type\_II

1. 0 0 0 0 1 66 | g = Obesity\_Type\_III