Obesity Analysis:
Analyzing Correlation & Predictive Models

Group B16

Yiwei Lu | Ziqi Zhang | Zaiheng Shen | Wan-Lun Tsai | Chia-Chien Chang

Underweight: BMI < 18.5

Normal weight: 18.5 ≤ BMI ≤ 24.9

Overweight: BMI between 25.0 and 29.9

- Overweight_Level_I: 25.0 ≤ BMI ≤ 27.4
- Overweight_Level_II: 27.5 ≤ BMI ≤ 29.9

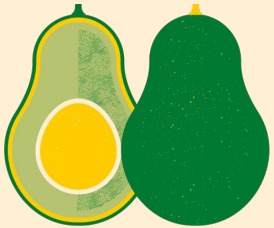
Obesity: BMI between 30 and 40+

- Obesity Type I: 30.0 ≤ BMI ≤ 34.9
- Obesity Type II: 35.0 ≤ BMI ≤ 39.9
- Obesity Type III: BMI ≥ 40

KEY QUESTION

1. Obesity Risk Prediction
2. Key Factors of Obesity
3. Impact of Habits on Obesity
4. Genetics and Family Influence

AGENDA



1. Exploratory Data Analysis

2. Data Engineering

- Decision Tree and Random Forest
- KNN
- Linear Regression

3. Conclusion

Exploratory Data Analysis

Data Understanding

Dataset : UCI Machine Learning Repository

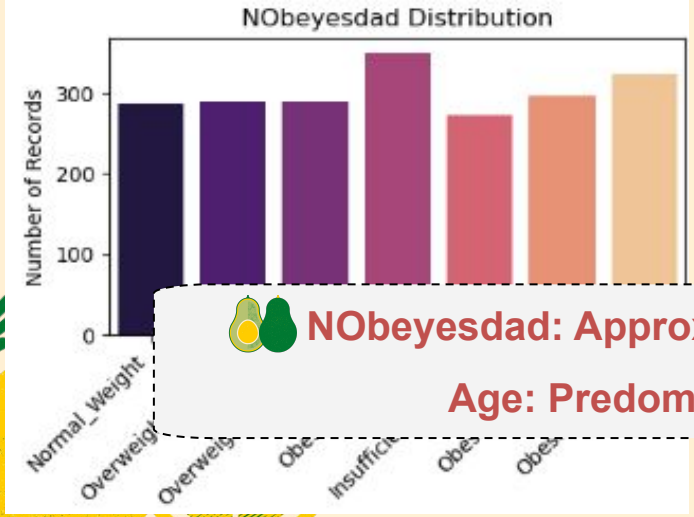
Data Shape : (2111, 17)

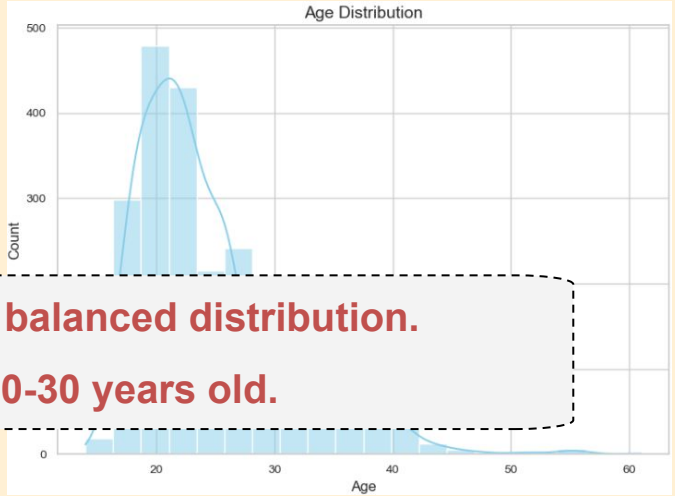
Target Variable: NObeyesdad

Missing Values : 0

Duplicated : 24

Outliers : (Age 168, NCP 579)





NObeyesdad: Approximately balanced distribution.

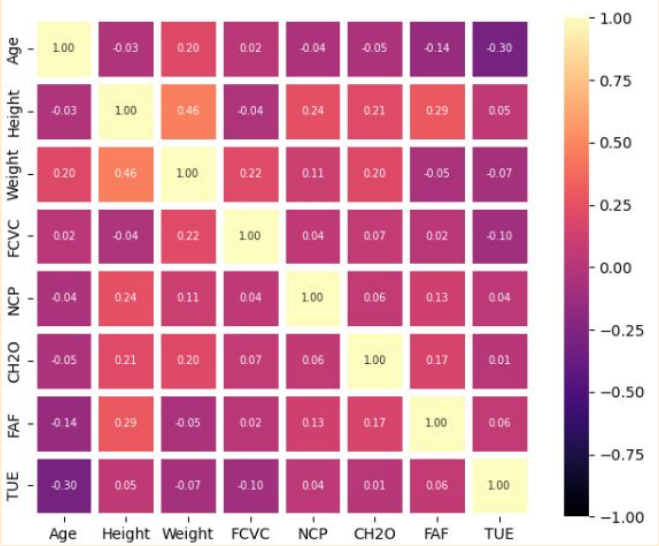
Age: Predominantly 20-30 years old.

Data Understanding

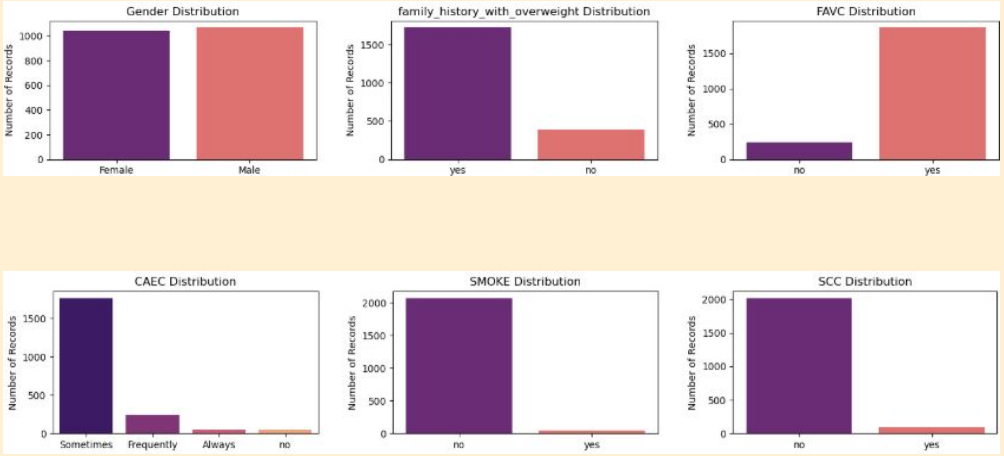
Attributes Related to Physical Condition	
SCC: Calories consumption monitoring	categorical
MTRANS: Transportation used	categorical
FAF: Physical activity frequency	numerical
TUE: Time using technology devices	numerical

Attributes Related to Eating Habits	
FAVC: Frequent consumption of high caloric food	categorical
CAEC: Consumption of food between meals	categorical
CALC: Consumption of alcohol	categorical
FCVC: Frequency of consumption of vegetables	numerical
NCP: Number of main meals	numerical
CH20: Consumption of water daily	numerical

Correlation b/w Numerical Variables



Categorical Variables Distribution



1. The correlation between CH2O and NCP is slightly higher (about 0.24)
2. FCVC(0.07 |vegetables consumption) and FAF(0.17| physical activity) have high correction with CH2O
3. There are 1,726 people who have a family history of being overweight.
4. 2,067 people are non-smokers, while 44 people smoke

Data Engineering

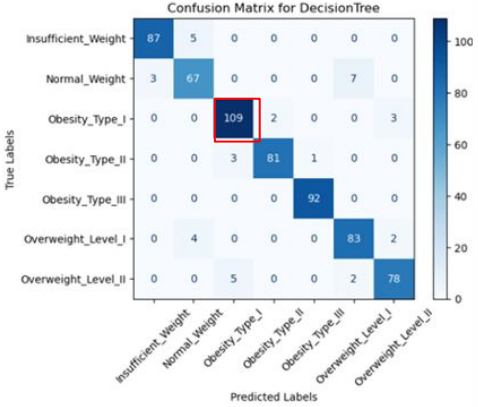
Obesity Risk Prediction

Model Performance Comparison

Decision Tree

```
Accuracy: 0.9416403785488959
Confusion Matrix:
[[ 87   5   0   0   0   0   0]
 [  3  67   0   0   0   7   0]
 [  0   0 109   2   0   0   3]
 [  0   0   3  81   1   0   0]
 [  0   0   0   0  92   0   0]
 [  0   4   0   0   0  83   2]
 [  0   0   5   0   0   2  78]]
Classification Report:
```

	precision	recall	f1-score	support
Insufficient_Weight	0.97	0.95	0.96	92
Normal_Weight	0.88	0.87	0.88	77
Obesity_Type_I	0.93	0.96	0.94	114
Obesity_Type_II	0.98	0.95	0.96	85
Obesity_Type_III	0.99	1.00	0.99	92
Overweight_Level_I	0.90	0.93	0.92	89
Overweight_Level_II	0.94	0.92	0.93	85
accuracy			0.94	634
macro avg	0.94	0.94	0.94	634
weighted avg	0.94	0.94	0.94	634

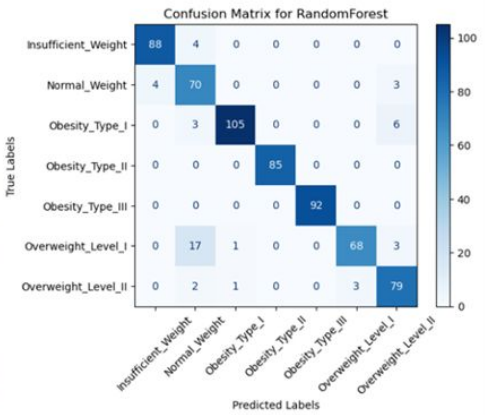


94.16%

Random Forest

```
Accuracy: 0.9258675078864353
Confusion Matrix:
[[ 88   4   0   0   0   0   0]
 [  4  70   0   0   0   0   3]
 [  0   3 105   0   0   0   6]
 [  0   0   0  85   0   0   0]
 [  0   0   0   0  92   0   0]
 [  0  17   1   0   0  68   3]
 [  0   2   1   0   0   3  79]]
Classification Report_RF:
```

	precision	recall	f1-score	support
Insufficient_Weight	0.96	0.96	0.96	92
Normal_Weight	0.73	0.91	0.81	77
Obesity_Type_I	0.98	0.92	0.95	114
Obesity_Type_II	1.00	1.00	1.00	85
Obesity_Type_III	1.00	1.00	1.00	92
Overweight_Level_I	0.96	0.76	0.85	89
Overweight_Level_II	0.87	0.93	0.90	85
accuracy			0.93	634
macro avg	0.93	0.93	0.92	634
weighted avg	0.93	0.93	0.93	634



92.58%

Obesity_Type_III: **99% accuracy**
Normal_Weight: **88% accuracy** (slightly higher error).

All categories have F1-scores **close to or above 90%**, indicating well-balanced performance.

Obesity_Type_II, Obesity_Type_III: **100% accuracy**.

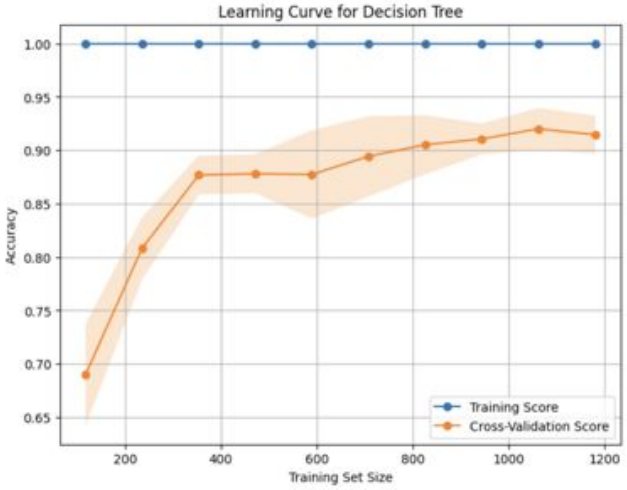
Normal_Weight: **73% accuracy** (some mispredictions); **91%recall**
Overweight_Level_I: **96% accuracy**; **76% recall** (some misclassified)

All categories have F1-scores **above 80%**, reflecting strong performance across most categories.

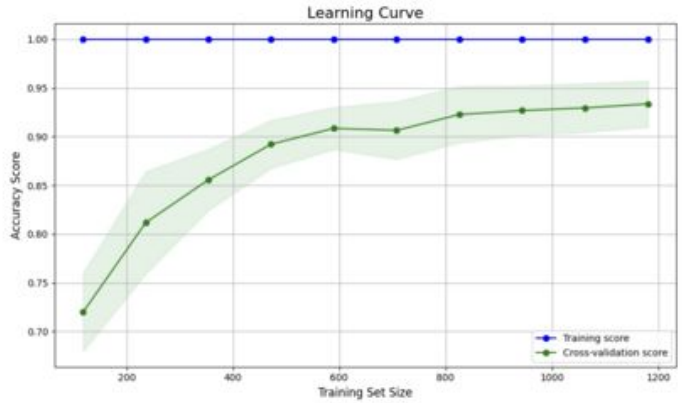
Obesity Risk Prediction

Model Performance Comparison

Decision Tree



Random Forest



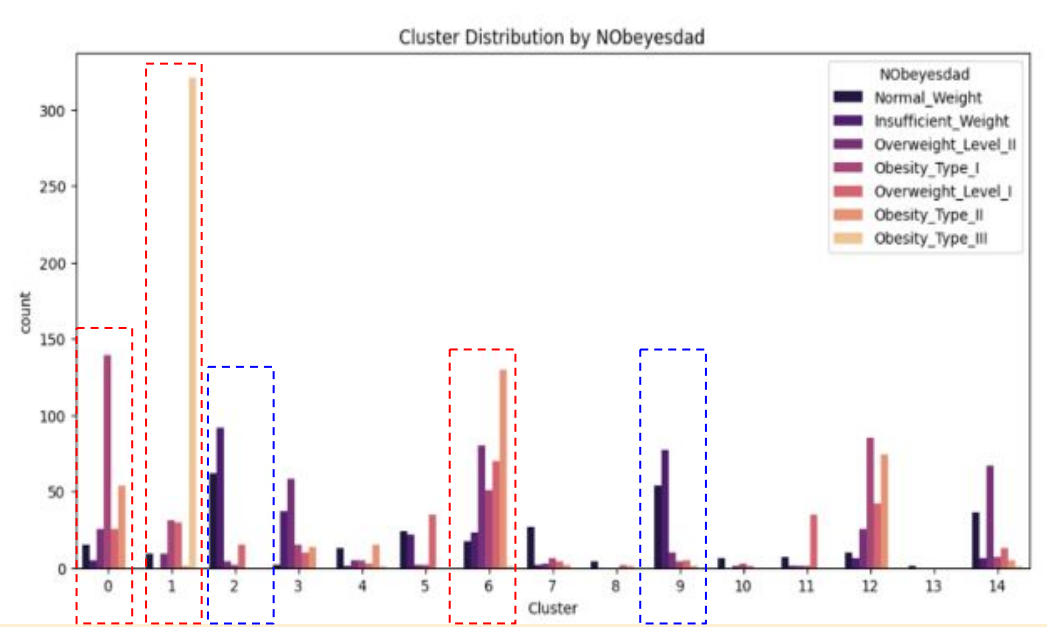
Decision Tree Accuracy: 94.16% | Random Forest Accuracy: 92.59%

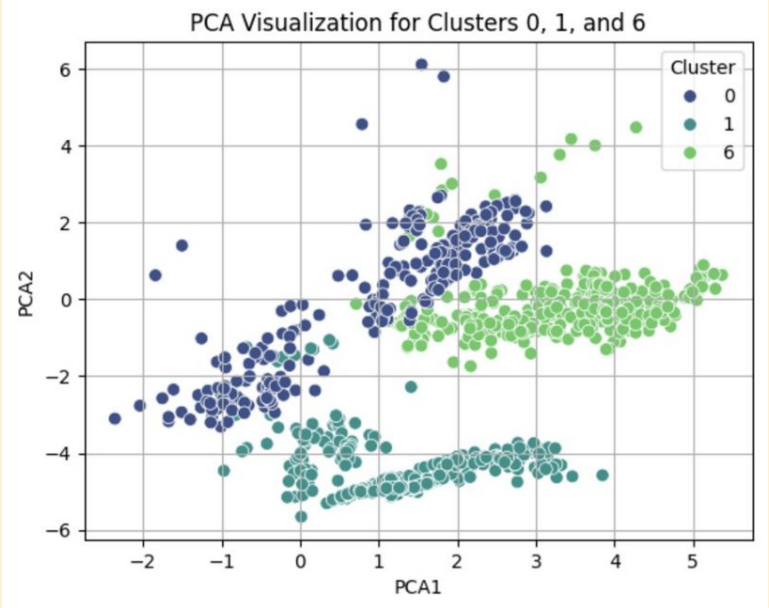
Both models show overfitting.

Random Forest is more stable compared to the Decision Tree.

Key Factors Influencing

(KNN Algorithm – 15 Clusters Analysis)





Obesity | Cluster 0,1,6

Non-Obesity | Cluster 2,9

Obesity

Common variables: ['Height', 'Weight', 'TUE', 'MTRANS_Walking', 'NCP', 'MTRANS_Public_Transportation', 'FCVC', 'Age', 'CH2O']

The key differences:

Cluster 0 (Obesity Type I | 60%)
▪Gender, No alcohol (CALC_no) and Frequent alcohol consumption (CALC_Frequently).

Cluster 1 (Obesity Type III | 90%)
▪Physical activity (FAF) BUT Drinking (CALC_Sometimes, frequently) habits influence severe obesity.

Cluster 6 (Obesity Type II | 44%)
▪High physical activity (FAF) BUT frequent snack intake (CAEC_Frequently, CAEC_Sometimes)

Non-Obesity

Common variables: ['Height', 'Weight', 'TUE', 'CALC_Frequently', 'MTRANS_Walking', 'NCP', 'MTRANS_Public_Transportation', 'FCVC', 'Age', 'CH2O"]

The key differences:

Cluster 2 (**normal/underweight | 88%)**

▪No alcohol consumption (**CALC_no**) contributes to lower caloric intake.

Cluster 9 (**normal/underweight | 86%)**

▪**High physical activity (FAF)** and occasional alcohol (**CALC_Sometimes**) maintain balance.

These factors key factor cause Obesity is:

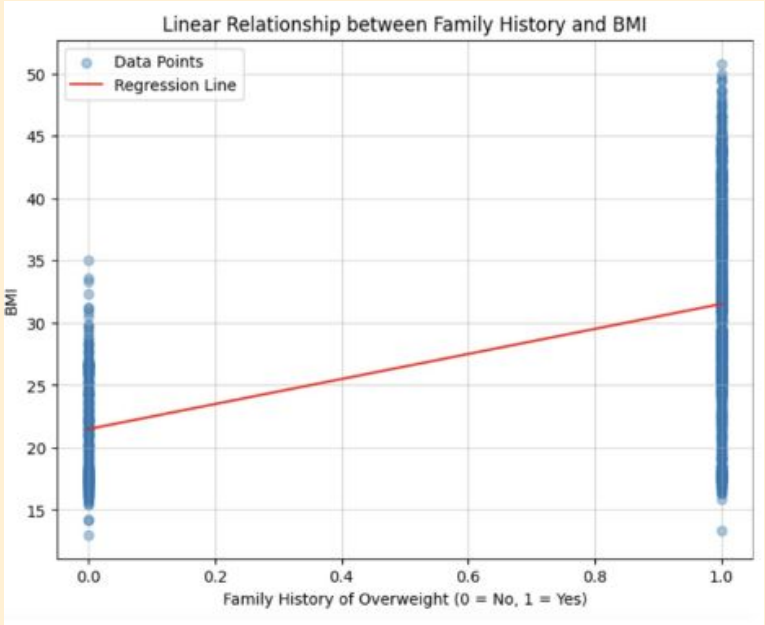
CALC_Sometimes, frequently; CAEC_Frequently, CAEC_Sometimes

Family History and Obesity

(Linear Regression Analysis)

P-Value: 0.000 (P < 0.05), indicating a statistically significant relationship.

Shows a positive linear relationship between family history and BMI.



Individuals with a family history of overweight tend to have higher BMI.

Conclusion

Prediction Model for Obesity Rick

Random Forest 92.59% Accuracy

Eating Habit

Frequent snack intake: CAEC_Frequently, CAEC_Sometimes
Alcohol consumption: CALC_Sometimes, frequently;

Physical Activity(FAF)

Regular exercise and meal patterns are crucial for maintaining a healthy weight.

Genetics

Family history has a significant impact on BMI.

Active lifestyles and avoiding alcohol support lower weight.
Irregular eating and frequent snacking contribute to obesity.

THANK YOU!