Sravan Phani Kumar Kopparthi
MSBD566 – Predictive Modeling and Analytics
10/08/2025

Assignment 2

Data Title	Estimation of Obesity Levels Based On Eating Habits and Physical Condition
Data Source	https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition

Data description: This dataset contains data for estimating obesity levels in individuals from Mexico, Peru, and Colombia, based on their eating habits and physical condition. The data contains 17 attributes and 2111 records, the records are labeled with the class variable NObesity (Obesity Level), which allows classification of the data using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III.

Data dictionary: all the definitions, what the columns mean, etc.

Column Descriptions:

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Gender	Feature	Categorical	Gender			no
Age	Feature	Continuous	Age			no
Height	Feature	Continuous				no
Weight	Feature	Continuous				no
family_history_with_overweight	Feature	Binary		Has a family member suffered or suffers from overweight?		no
FAVC	Feature	Binary		Do you eat high caloric food frequently?		no

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
FCVC	Feature	Integer		Do you usually eat vegetables in your meals?		no
NCP	Feature	Continuous		How many main meals do you have daily?		no
CAEC	Feature	Categorical		Do you eat any food between meals?		no
SMOKE	Feature	Binary		Do you smoke?		no
CH2O	Feature	Continuous		How much water do you drink daily?		no
SCC	Feature	Binary		Do you monitor the calories you eat daily?		no
FAF	Feature	Continuous		How often do you have physical activity?		no
TUE	Feature	Integer		How much time do you use technological devices such as cell phone, videogames, television, computer and others?		no
CALC	Feature	Categorical		How often do you drink alcohol?		no
MTRANS	Feature	Categorical		Which transportation do you usually use?		no
NObeyesdad	Target	Categorical		Obesity level		no

Datatypes for Variables:

```
Gender      object
Age        float64
Height     float64
Weight     float64
```

family_history_with_overweight object
FAVC object
FCVC float64
NCP float64
CAEC object
SMOKE object
CH2O float64
SCC object
FAF float64
TUE float64
CALC object
MTRANS object
NObeyesdad object

Describe of the Dataset

	Age	Height	Weight	FCVC	NCP	CH2O	FAF	TUE
count	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000
mean	24.312600	1.701677	86.586058	2.419043	2.685628	2.008011	1.010298	0.657866
std	6.345968	0.093305	26.191172	0.533927	0.778039	0.612953	0.850592	0.608927
min	14.000000	1.450000	39.000000	1.000000	1.000000	1.000000	0.000000	0.000000

	Age	Height	Weight	FCVC	NCP	CH2O	FAF	TUE
25%	19.947192	1.630000	65.473343	2.000000	2.658738	1.584812	0.124505	0.000000
50%	22.777890	1.700499	83.000000	2.385502	3.000000	2.000000	1.000000	0.625350
75%	26.000000	1.768464	107.430682	3.000000	3.000000	2.477420	1.666678	1.000000
max	61.000000	1.980000	173.000000	3.000000	4.000000	3.000000	3.000000	2.000000

No Null values in the dataset.

Disclaimer:

77% of the data was generated synthetically using the Weka tool and the SMOTE filter, 23% of the data was collected directly from users through a web platform.