OBESITY PREDICTION

By Gracie Inman

OVERVIEW OF TOPIC

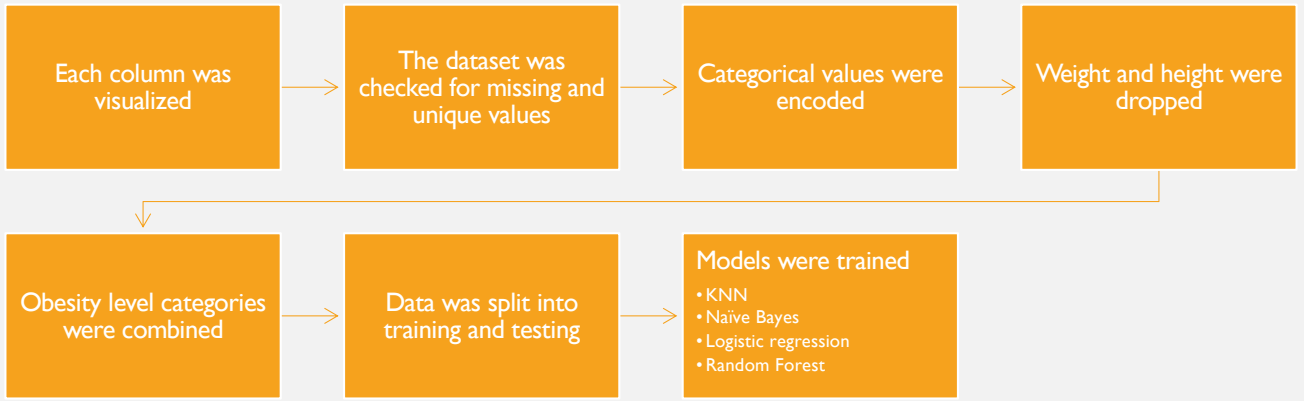
- Obesity is a disease that affected 1 out of 8 people in 2022 (World Health Organization, 2024).
- According to the World Health Organization (2024) , obesity can also cause other negative health impacts such as:
  - Increased risk of type 2 diabetes and heart disease
  - Negatively impact bone health and reproduction
  - Increased risk of certain cancers

DATA SET AND VARIABLES

- The Obesity levels dataset was obtained from Kaggle (Mehrparvar 2024).
- The obesity levels dataset contains numerous variables regarding the lifestyle of the individual and their obesity level.
- The dataset contains 77% simulated and 23% web-obtained data.

- Gender: Feature, Categorical, "Gender"
- Age: Feature, Continuous, "Age"
- Height: Feature, Continuous
- Weight: Feature Continuous
- family_history_with_overweight: Feature, Binary, " Has a family member suffered or suffers from overweight? "
- FAVC: Feature, Binary, " Do you eat high caloric food frequently? "
- FCVC: Feature, Integer, " Do you usually eat vegetables in your meals? "
- NCP: Feature, Continuous, " How many main meals do you have daily? "
- CAEC: Feature, Categorical, " Do you eat any food between meals? "
- SMOKE: Feature, Binary, " Do you smoke? "
- CH2O: Feature, Continuous, " How much water do you drink daily? "
- SCC: Feature, Binary, " Do you monitor the calories you eat daily? "
- FAF: Feature, Continuous, " How often do you have physical activity? "
- TUE: Feature, Integer, " How much time do you use technological devices such as cell phones, video games, television, computers, and others? "
- CALC: Feature, Categorical, " How often do you drink alcohol? "
- MTRANS: Feature, Categorical, " Which transportation do you usually use? "
- NObeyesdad: Target, Categorical, "Obesity level"

METHODS


Each column was visualized
The dataset was checked for missing and unique values
Categorical values were encoded
Weight and height were dropped
Obesity level categories were combined
Data was split into training and testing
Models were trained
• KNN
• Naïve Bayes
• Logistic regression
• Random Forest

RESULTS

- Shown to the right are the evaluation metrics for each model

Table 1: KNN

	Precision	Recall	F1-score	Support
0	0.68	0.91	0.78	56
1	0.63	0.19	0.30	62
2	0.80	0.96	0.87	199
3	0.81	0.67	0.73	106
Accuracy			0.77	423
Macro avg	0.73	0.68	0.67	423
Weighted avg	0.76	0.77	0.74	423

Table 2: Naïve Bayes

	Precision	Recall	F1-score	Support
0	0.39	0.66	0.49	56
1	0.42	0.18	0.25	62
2	0.71	0.94	0.81	199
3	0.58	0.21	0.31	106
Accuracy			0.61	423
Macro avg	0.52	0.50	0.46	423
Weighted avg	0.59	0.61	0.56	423

Table 3: Logistic Regression

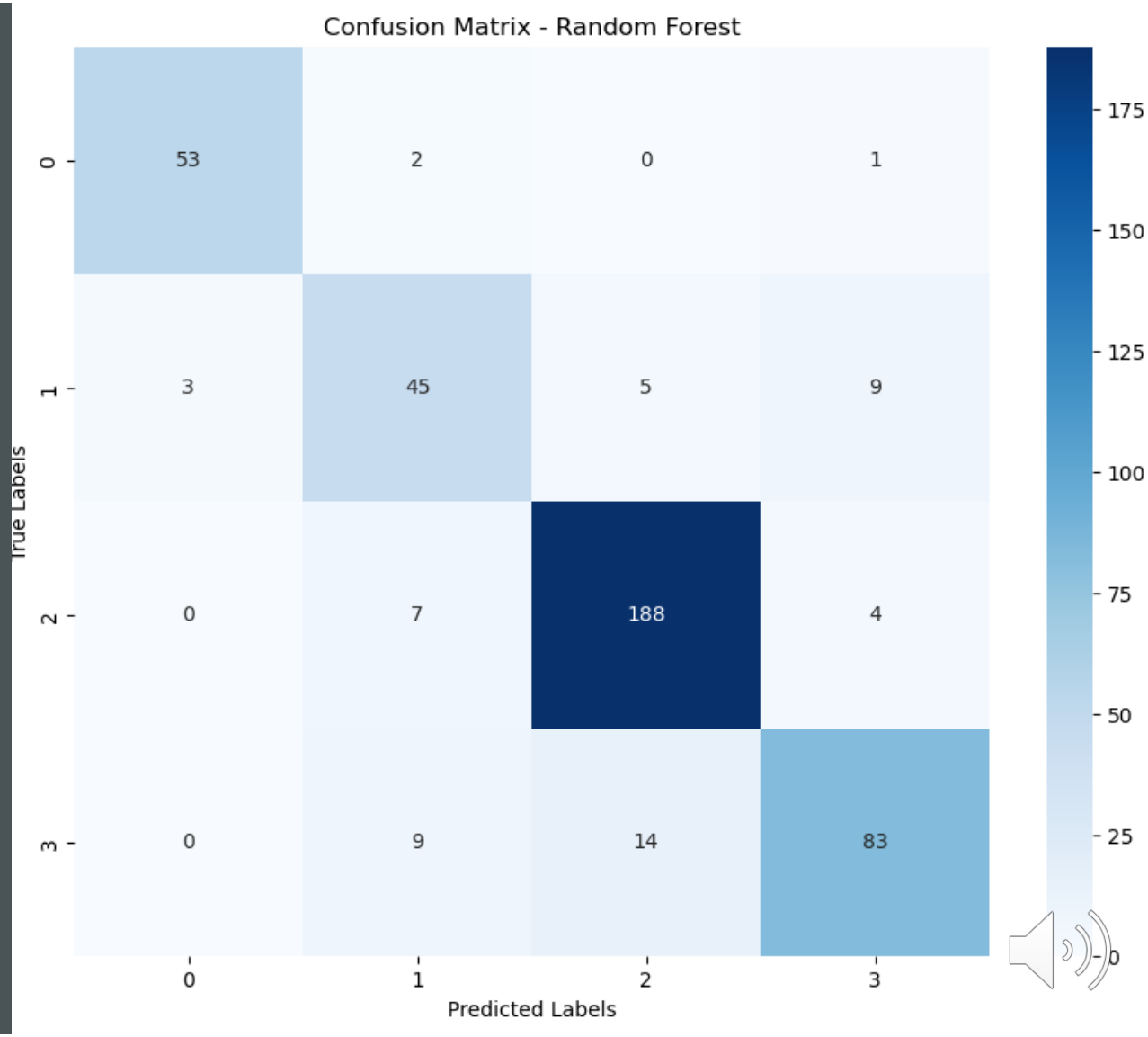
	Precision	Recall	F1-score	Support
0	0.52	0.54	0.53	56
1	0.50	0.24	0.33	62
2	0.72	0.90	0.80	199
3	0.52	0.42	0.47	106
Accuracy			0.64	423
Macro avg	0.56	0.53	0.53	423
Weighted avg	0.61	0.64	0.61	423

Table 4: Random Forest

	Precision	Recall	F1-score	Support
0	0.95	0.95	0.95	56
1	0.71	0.73	0.72	62
2	0.91	0.94	0.93	199
3	0.86	0.78	0.82	106
Accuracy			0.87	423
Macro avg	0.86	0.85	0.85	423
Weighted avg	0.87	0.87	0.87	423

CONFUSION MATRIX

- Shown to the right is the confusion matrix for the Random Forest Model



SUMMARY AND IMPORTANT CONSIDERATIONS

- Random Forest performed the best.
- Limitations
- Assumptions
- Ethical concerns
- Questions?

REFERENCES

- Centers for Disease Control and Prevention. (2022, May 17). *Adult obesity facts*. Centers for Disease Control and Prevention. https://www.cdc.gov/obesity/data/adult.html
- Mehrparvar, F. (2024, April 7). *Obesity levels*. Kaggle. https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels
- World Health Organization. (2024, March 1). *Obesity and overweight*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight#:~:text=Overview,the%20risk%20of%20certain%20cancers.