New Year, New You: Using Machine Learning to Predict Risk Factors of Obesity

PRESENTED BY: TASHA SANTIAGO, ALFRED SAVIO, CONNOR CHADWICK, WISDOM MBALUKA, AND CALUM HEWSON

01.14.25
PROJECT 4

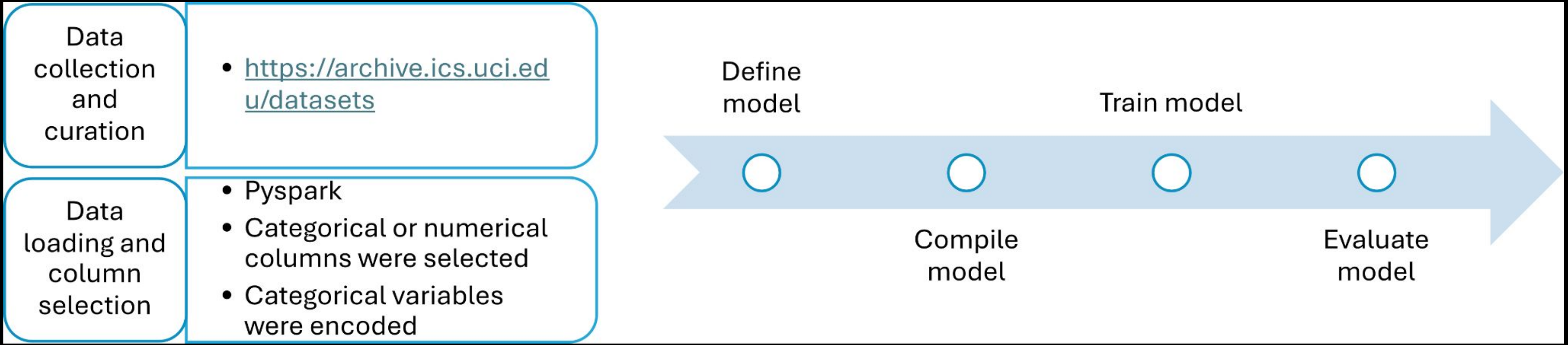
Predicting the factors associated with the risk of obesity is crucial

- **Early Identification**: By identifying individuals at risk early, we can intervene before obesity-related complications develop, leading to better long-term health outcomes

- **Prevention**: Understanding the underlying risk factors helps in designing targeted prevention strategies, potentially reducing the prevalence of obesity and its associated health risks

- **Informed Healthcare Strategies**: Accurate predictions allow for more personalized health advice, lifestyle recommendations, and treatment plans, improving overall care

- **Global Health Impact**: With 1 in 8 people globally living with obesity (2022) and rates expected to rise, predicting these risk factors is vital to addressing the growing obesity epidemic and mitigating its economic and healthcare burdens



NEURAL NETWORK PROCESS OVERVIEW

- Model was developed, compiled and evaluated using each numerical variable:
    - Height
    - Age
    - Meals per day
    - Frequency physical activity
    - Water intake
    - Vegetable intake
    - Technology use
- Results showed that ‘Age’ and ‘Meals per day’ had the highest accuracies in predicting the risk of obesity

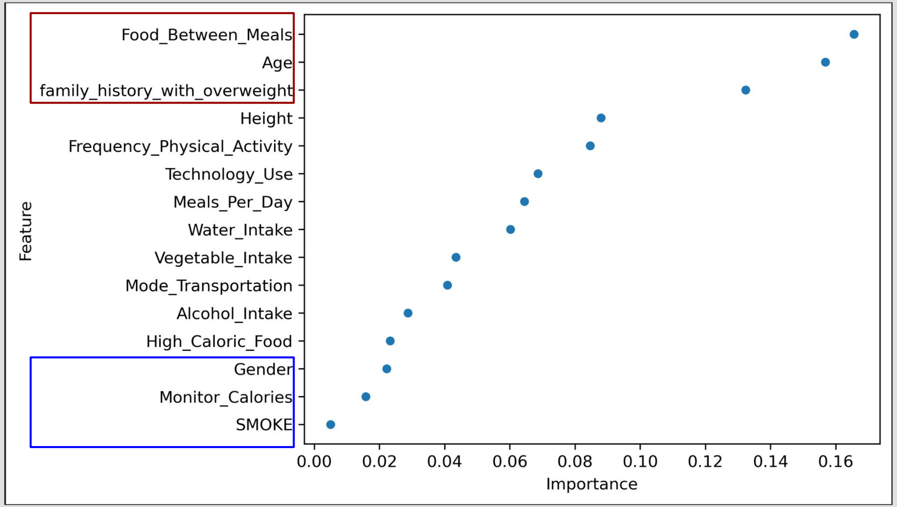
‘Age’ & ‘Meals per day’ showed the highest accuracy from numerical variables

Numerical Variables	Neural Network Accuracy
Meals per day	**77.8%**
Age	**75.9%**
Height	73.5%
Water intake	73.5%
Vegetable intake	73.5%
Technology use	73.5%
Frequency physical activity	72.1%

‘Food between meals’ and ‘Family history’ showed the highest accuracy from categorical variables

- Model was developed, compiled and evaluated using each categorical variable:
  - Food between meals
  - Mode of transportation
  - Alcohol intake
  - Smoking
  - Family history
  - High-caloric food intake
  - Monitoring calories
  - Gender
- Results showed that ‘Food between meals’, ‘Family history’ and ‘Monitoring calories’ (borderline) had the highest accuracies in predicting the risk of obesity

Categorical Variables	Neural Network Accuracy
Food between meals	**83.7%**
Family history	**83.7%**
Monitoring calories	**74.9%**
Mode of transportation	73.8%
Alcohol intake	73.5%
Smoking	73.5%
High-caloric food intake	73.5%
Gender	73.5%



	Feature	Importance
10	SMOKE	0.005017
13	Monitor_Calories	0.015747
14	Gender	0.022262
12	High_Caloric_Food	0.023310
9	Alcohol_Intake	0.028763
8	Mode_Transportation	0.040846
5	Vegetable_Intake	0.043423
4	Water_Intake	0.060245
2	Meals_Per_Day	0.064420
6	Technology_Use	0.068598
3	Frequency_Physical_Activity	0.084680
0	Height	0.087969
11	family_history_with_overweight	0.132402
1	Age	0.156754
7	Food_Between_Meals	0.165562

So... which are the most important categories?

- A Random Forest model was trained using RandomForestClassifier()
- Features were ranked in order of importance using rf_model.feature_importances_
- ‘Food between meals’, ‘Age’ and ‘Family history’ are the top three factors influencing/predicting risk of obesity

Conclusions

- Model Overview:
  - Deep Learning (TensorFlow) and Random Forest Classifier were used to predict Obesity/Overweight based on demographic and lifestyle features
- Best Predictors:
  - Categorical Variables:
    - ‘Food Between Meals’ and ‘Family History of Overweight’: 83.7% accuracy—strong predictors
  - Numerical Variables:
    - ‘Meals Per Day’: 77.8% accuracy—important for prediction
  - Moderate Predictive Power:
    - Other features (e.g., Age, Physical activity, Technology use) showed moderate accuracy (~72-77%), indicating room for further improvement
  - Model Performance:
    - The loss values (e.g., Age, Food Between Meals) indicate overall good model performance, with areas for refinement, especially for variables with higher loss

Conclusions (cont.)

IMPROVEMENT IDEAS:

Feature Engineering:

- *Explore additional relevant features* such as sleep quality, stress levels, mental health factors, and genetic predispositions, which are crucial but not yet considered
- *Transform features* (e.g., interactions between physical activity frequency and meals per day) to capture more complex relationships within the data

Data Expansion:

- *Increase the sample size* to improve model generalization and robustness, especially for underrepresented subgroups or minority classes
- *Augment the dataset* by collecting more real-world data from diverse demographics and environments

Acknowledgements

- Fellow students
- TAs
- Sub-Instructors
- Reed Hyde
- Anish Talsania

QUESTIONS?

New Year, New You: Using Machine Learning to Predict Risk Factors of Obesity

01.14.25
PROJECT 4

PRESENTED BY: TASHA SANTIAGO, ALFRED SAVIO, CONNOR CHADWICK, WISDOM MBALUKA, AND CALUM HEWSON

“ Predicting the factors associated with the risk of obesity is crucial

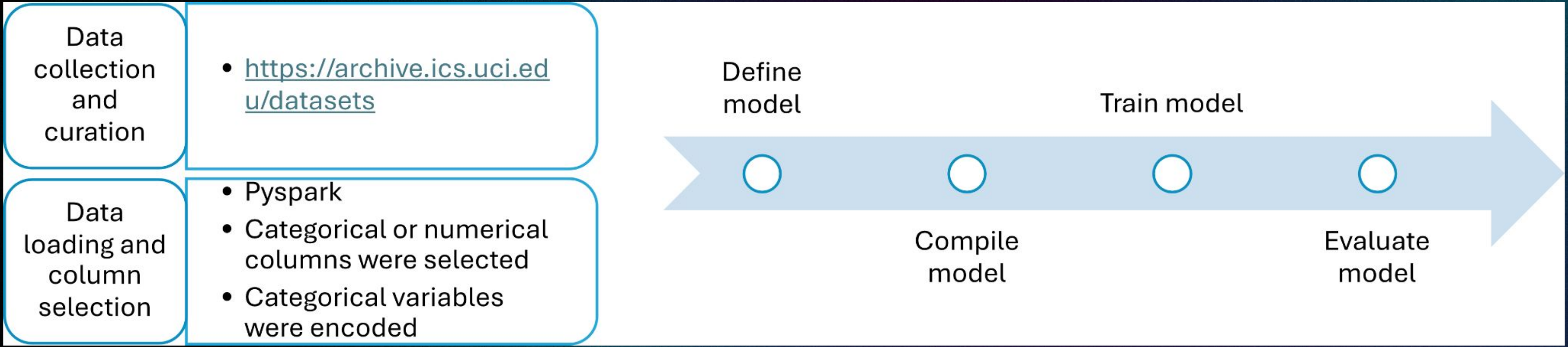
- **Early Identification**: By identifying individuals at risk early, we can intervene before obesity-related complications develop, leading to better long-term health outcomes

- **Prevention**: Understanding the underlying risk factors helps in designing targeted prevention strategies, potentially reducing the prevalence of obesity and its associated health risks

- **Informed Healthcare Strategies**: Accurate predictions allow for more personalized health advice, lifestyle recommendations, and treatment plans, improving overall care

- **Global Health Impact**: With 1 in 8 people globally living with obesity(2022) and rates expected to rise, predicting these risk factors is vital to addressing the growing obesity epidemic and mitigating its economic and healthcare burdens

Neural Network Process Overview



“ ‘Age’ and ‘Meals per day’ showed the highest accuracy from numerical variables

- Model was developed, compiled and evaluated using each numerical variable:
  - Height
  - Age
  - Meals per day
  - Frequency physical activity
  - Water intake
  - Vegetable intake
  - Technology use
- Results showed that ‘Age’ and ‘Meals per day’ had the highest accuracies in predicting the risk of obesity

Numerical Variables	**Neural Network Accuracy**
Meals per day	**77.8%**
Age	**75.9%**
Height	73.5%
Water intake	73.5%
Vegetable intake	73.5%
Technology use	73.5%
Frequency physical activity	72.1%

“ ‘Food between meals’ and ‘Family history’ showed the highest accuracy from categorical variables

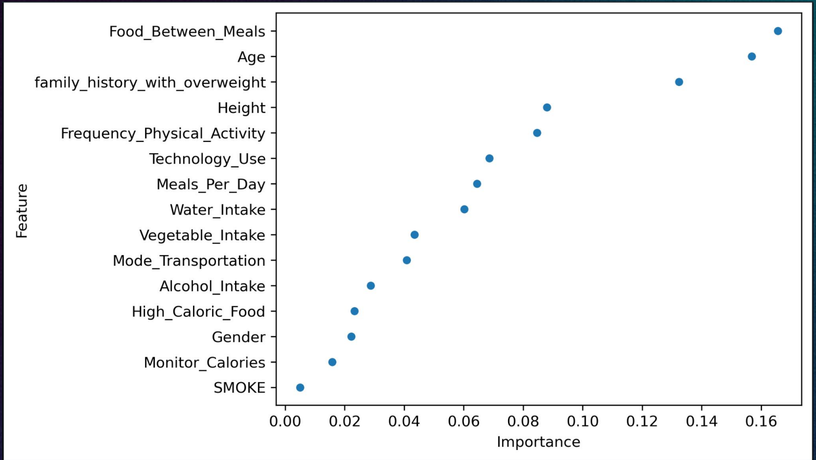
- Model was developed, compiled and evaluated using each categorical variable:
  - Food between meals
  - Mode of transportation
  - Alcohol intake
  - Smoking
  - Family history
  - High-caloric food intake
  - Monitoring calories
  - Gender
- Results showed that ‘Food between meals’, ‘Family history’ and ‘Monitoring calories’ (borderline) had the highest accuracies in predicting the risk of obesity

Categorical Variables	Neural Network Accuracy
Food between meals	**83.7%**
Family history	**83.7%**
Monitoring calories	**74.9%**
Mode of transportation	73.8%
Alcohol intake	73.5%
Smoking	73.5%
High-caloric food intake	73.5%
Gender	73.5%

So… which are the most important factors?

“

- A Random Forest model was trained using RandomForestClassifier()
- Features were ranked in order of importance using rf_model.feature_importances_
- ‘Food between meals’, ‘Age’ and ‘Family history’ are the top three factors influencing/predicting risk of obesity



	Feature	Importance
10	SMOKE	0.005017
13	Monitor_Calories	0.015747
14	Gender	0.022262
12	High_Caloric_Food	0.023310
9	Alcohol_Intake	0.028763
8	Mode_Transportation	0.040846
5	Vegetable_Intake	0.043423
4	Water_Intake	0.060245
2	Meals_Per_Day	0.064420
6	Technology_Use	0.068598
3	Frequency_Physical_Activity	0.084680
0	Height	0.087969
11	family_history_with_overweight	0.132402
1	Age	0.156754
7	Food_Between_Meals	0.165562

Conclusions

“

- **Model Overview**:
  - **Deep Learning (TensorFlow)** and **Random Forest Classifier** were used to predict **Obesity/Overweight** based on demographic and lifestyle features
- **Best Predictors**:
  - **Categorical Variables**:
    - **‘Food Between Meals’** and ‘**Family History of Overweight’**: **83.7%** accuracy—strong predictors
  - **Numerical Variables**:
    - **‘Meals Per Day’**: **77.8%** accuracy—important for prediction

Conclusions (cont.)

“

- **Moderate Predictive Power**:
  - Other features (e.g., **Age**, **Physical activity**, **Technology use**) showed moderate accuracy (~72-77%), indicating room for further improvement
- **Model Performance**:
  - The **loss values** (e.g., **Age**, **Food Between Meals**) indicate overall good model performance, with areas for refinement, especially for variables with higher loss