# Obesity Risk Prediction in Adults using Machine Learning

Submitted By: Ayush Bhatnagar
CWID ID: 885491738

Submitted To: Professor Kanika Sood
Date: 12/13/2024

# Introduction

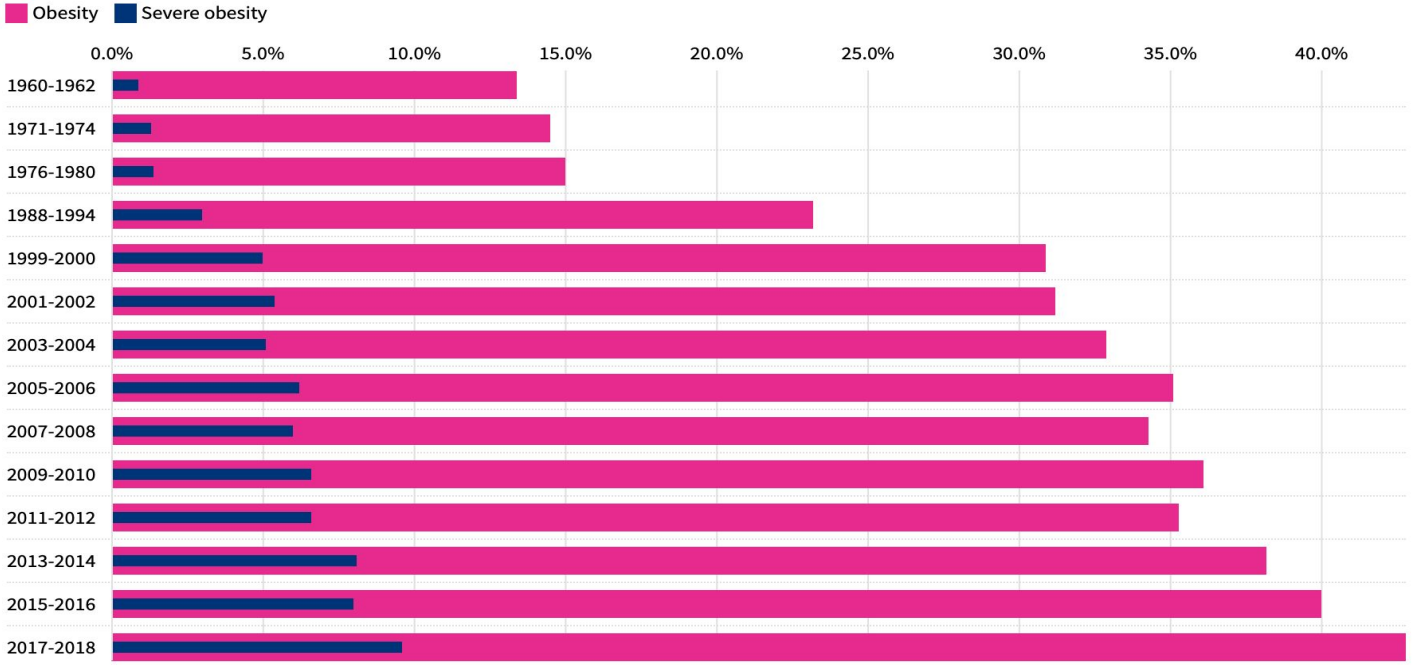
**Understanding Obesity:**
Obesity is a global health crisis linked to chronic illnesses like diabetes and heart disease. With nearly **70%** of U.S. adults overweight or obese, it remains a leading **preventable** cause of mortality.

**Need for Predictive Analysis:**
Preventing obesity requires not only lifestyle interventions but also predictive tools to identify at-risk populations early. **Predictive analysis** offers a strategic advantage by leveraging data to guide public health initiatives and personal health decisions, making prevention more effective and resource-efficient.

**Machine Learning as a Catalyst:**
Machine learning enables advanced analysis of complex data to uncover patterns and risk factors for obesity. It enhances prediction accuracy, supporting early interventions and personalized healthcare strategies.

# Statistics

Nationwide obesity rates have more than **tripled** since the 1960s !




*Courtesy-USA Facts

# Problem Statement

The global **obesity epidemic** is a growing concern, highlighting the urgent need for better **predictive analysis** and proactive **intervention strategies**. Despite the transformative potential of **machine learning** in identifying **obesity risk factors**, several challenges, such as **data quality, feature variability,** and **model generalizability**, have limited its widespread adoption.

This project seeks to overcome these obstacles by leveraging advanced **machine learning techniques** to build more precise **predictive models**. These models aim to not only identify individuals at risk of obesity but also guide the development of targeted **public health interventions,** contributing to a more effective response to this pressing issue.

# Objective

**Ultimate Goal:**

The objective is to detect individuals at an elevated risk of developing obesity at an early stage, facilitating timely intervention and preventive actions.

**Solutions Provided:**

- Leverage machine learning models and techniques to enhance the accuracy of obesity risk prediction.
- Offering insights into how factors like diet and drinking habits contribute to obesity risk.
- Identifying key lifestyle behaviors that could be targeted for preventive interventions.

# Data Description

**Dataset Acquisition:**
Extensive online research was performed to locate datasets from health surveys and records. A suitable dataset was identified on OpenML, providing a broad range of features for analysis.

**Key Features:**
Additional variables such as Age Group, BMI, and Weight-to-Normal Ratio were incorporated to enhance prediction accuracy and uncover deeper insights into obesity risk. These features were selected for their relevance to lifestyle and health patterns.

**Data Preparation:**
The dataset underwent thorough cleaning, including addressing missing values, standardizing units, and ensuring consistency among variables. Outliers were managed effectively to guarantee data reliability, paving the way for accurate analysis and machine learning model training.

# Methodology

**Feature Categorization:**
The features were divided into two groups: numerical ones like Age and Weight, which required minimal processing, and categorical ones like Gender and Family History, which needed transformation to be model-ready.

**Data Scaling and Encoding:**
Numerical features were scaled with tools like StandardScaler for consistency, while categorical features were encoded using methods such as OneHotEncoder to make them suitable for the model.

**Target Labeling:**
The target variable, 'NObeyesdad,' was transformed into numerical labels to facilitate efficient processing and understanding by the model.

**Pipeline Integration:**
A streamlined pipeline was developed, combining preprocessing techniques with the LGBM classifier to ensure efficient and effective training.

# Machine Learning Pipeline

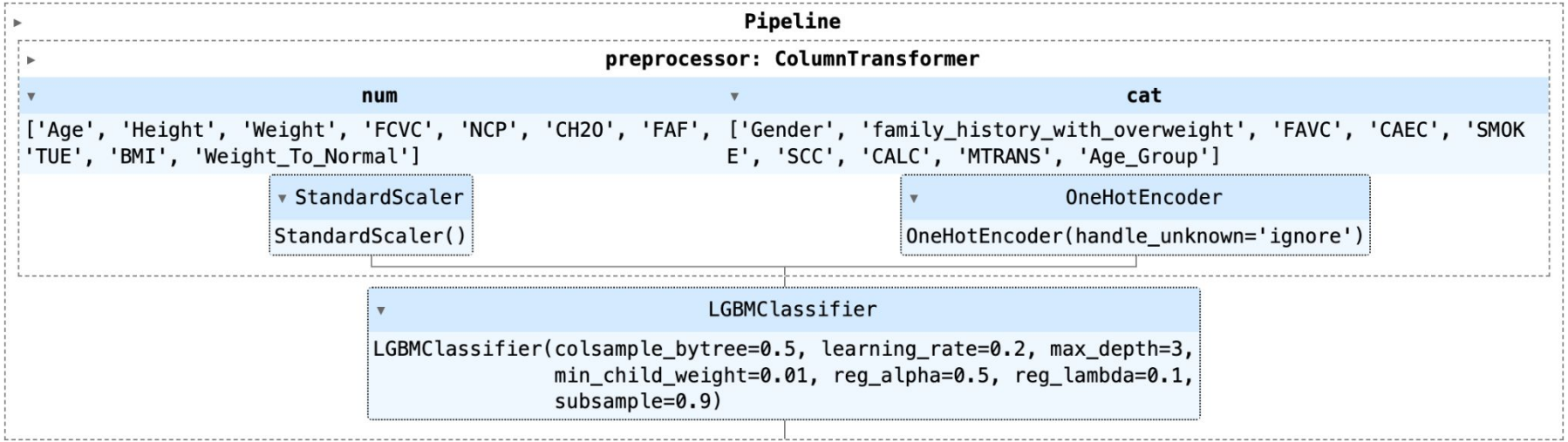
**Selected Algorithm**:
LightGBM (Light Gradient Boosting Machine), an enhanced version of Gradient Boosting Decision Trees (GBDT), was employed for its superior efficiency and scalability. As a supervised learning method, it excels in speed and performance compared to traditional GBDT models.

Advantages of LightGBM:

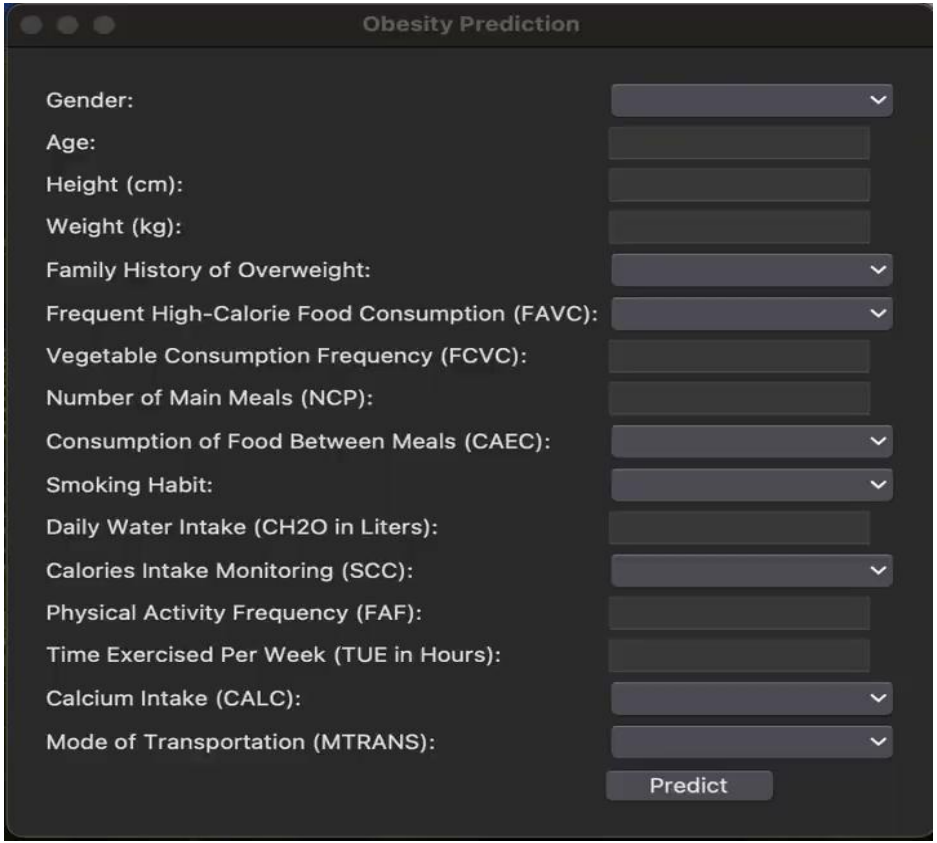
**High Efficiency and Robust Performance**: Provides rapid training and can handle large-scale datasets with ease. It effectively reduces overfitting while delivering excellent accuracy, even in tasks with imbalanced data.

**Rationale for Choosing LightGBM**:
LightGBM stood out due to its outstanding performance in classification problems, offering higher accuracy than many alternatives. Its capability to manage imbalanced datasets reliably makes it a strong contender for this project's predictive goals.

# Pipeline Architecture



# Demonstration of Results and Model Performance



# Key Insights

Even small factors, such as daily habits, diet, and activity level, can significantly impact BMI predictions. These subtle influences determine how much weight a person needs to lose or gain to reach a healthy, normal weight.

Understanding these factors allows for more personalized and effective weight management strategies.

# Challenges and Limitations

**Selecting Relevant Features:**
Identifying the most impactful features was a key hurdle. A detailed exploratory data analysis (EDA) and feature engineering were employed to uncover the variables that would provide the most reliable predictions.

**Tackling Model Overfitting:**
Overfitting posed a risk to the model's generalization. To address this, regularization methods such as L2 regularization and cross-validation were used to ensure consistent model performance across both training and unseen data.

**Communicating Model Outcomes:**
Effectively conveying the model's predictions was crucial. Clear and informative visualizations were designed to simplify the interpretation of results, making them more accessible and actionable for users.

# Conclusion

**Impact of the Project:**

This project plays a crucial role in the early detection of obesity, offering a dependable tool for assessing the risk of developing the condition. With this approach, healthcare providers can develop more targeted treatment plans and individuals can be encouraged to adopt healthier lifestyle habits.

**Future Directions:**

Moving forward, the emphasis will be on enhancing the model’s performance through further testing and optimization. Additionally, the next steps involve preparing for the model's deployment, including creating a user-friendly platform that makes the tool accessible to a wider range of users.

# Thank You For Watching!

Any Questions?