Predicting Adult Weight Using Machine Learning

Executive Insights for Scalable Healthcare Analytics

Thandiwe Moyo

14/12/2025

Agenda

1. Introduction to Weight Prediction
2. Data Preparation and Feature Selection
3. Modeling and Performance Evaluation
4. Implementation and Insights

Introduction to Weight Prediction

Thandiwe Moyo

Transforming Weight Prediction

Business Value of Weight Prediction

Enables targeted healthcare, insurance, and wellness solutions

Scalable, Data-Driven Insights

Machine learning delivers accurate, actionable predictions at scale

Applied Machine Learning Expertise

Demonstrates real-world impact and decision-making support

Business Impact & Relevance

Personalized Health Insights

Accurate weight prediction enables tailored care and risk profiling for individuals.

Proactive Healthcare Interventions

Early identification of at-risk adults supports timely, targeted health actions.

Operational Efficiency & Cost Savings

Optimized resource allocation reduces unnecessary expenditures for healthcare providers.

Defining the Challenge

Business Objective

Predict adult weight using a dataset with 617 variables.

Key Data Challenges

High dimensionality, missing values, and multicollinearity required advanced handling.

Solution Approach

Applied robust, scalable machine learning to ensure reliable, actionable results.

Data Preparation and Feature Selection

Thandiwe Moyo

Strategic Data Preparation

Streamlined Data for Clarity

Eliminated identifier and all-null columns to focus on relevant predictors.

Robust Handling of Missing Values

Imputed missing data using medians, ensuring resilience to outliers.

Consistent Feature Scaling

Applied RobustScaler to standardize features, minimizing the impact of anomalies.

Smart Feature Selection

Automated Feature Importance

Utilized Random Forest and XGBoost to rank variables by predictive power.

Targeted Variable Reduction

Selected top predictors to minimize model complexity and risk of overfitting.

Efficiency and Accuracy Balance

Streamlined input features while maintaining high predictive accuracy.

Modeling and Performance Evaluation

Thandiwe Moyo

Modeling Approach Overview

Comprehensive Model Evaluation

Assessed Linear, Ridge, Lasso, Random Forest, and XGBoost models for predictive accuracy

Rigorous Hyperparameter Tuning

Employed GridSearchCV with 5-fold cross-validation to optimize model performance

Focus on Real-World Impact

Prioritized models that generalize well and deliver reliable results on unseen data

Model Comparison at a Glance

Model	MSE	R^2
XGBoost	211.76	0.86
Random Forest	217.35	0.86
Linear Regression	476.19	0.69
Ridge Regression	467.42	0.70
Lasso Regression	467.39	0.70

Performance Metrics Table

Model	MSE	R^2
Linear Regression	476	0.69
Ridge Regression	467.4	0.70
Lasso Regression	467.38	0.70
Random Forest	217.35	0.86
XGBoost	211.76	0.86

Why XGBoost Wins

Best-in-Class Predictive Accuracy

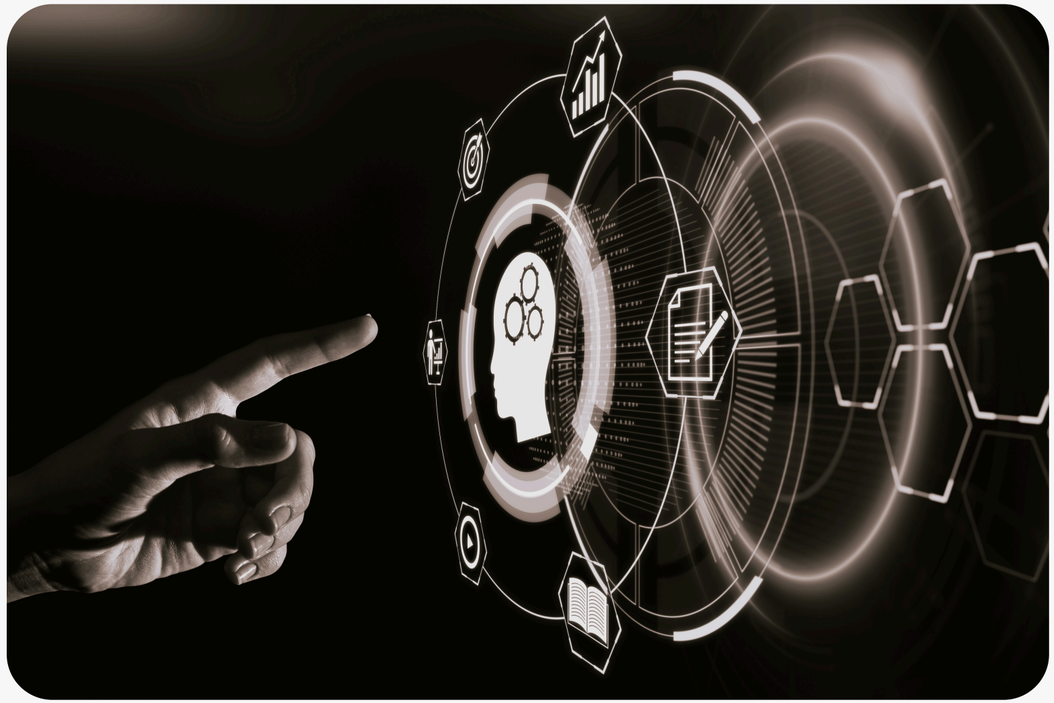
XGBoost achieved the lowest MSE (211.76) and highest R^2 (0.86) on test data.

Superior Handling of Complex Data

Excels with high-dimensional, non-linear, and correlated features.

Optimized for Efficiency

Feature selection and hyperparameter tuning reduce overfitting and boost generalizability.



Implementation and Insights

Thandiwe Moyo

Scalability & Decision-Making

Robust, Scalable Solution

Model architecture supports deployment across diverse healthcare environments.

Data-Driven Decision Support

Enables informed risk assessment, resource allocation, and care management.

Adaptable Analytics Framework

Approach can be extended to other predictive healthcare and business challenges.

Applied ML Expertise in Action

End-to-End ML Workflow

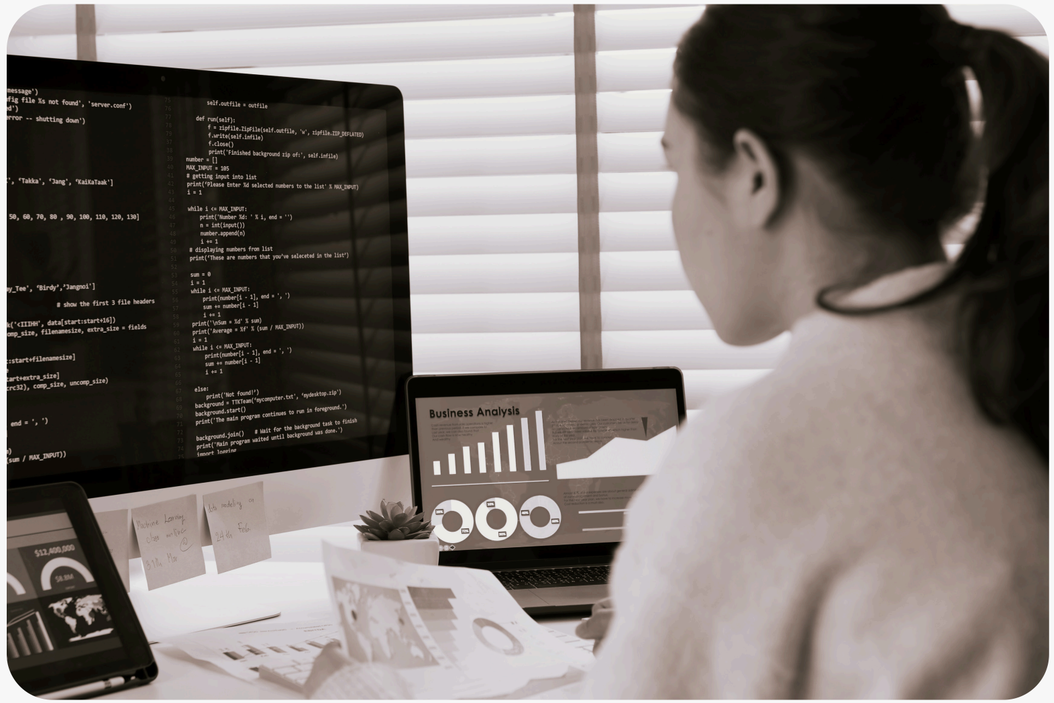
Executed data cleaning, robust feature selection, and advanced model validation.

Balanced Model Performance

Prioritized accuracy, efficiency, and generalizability for scalable, real-world deployment.

Strategic Technical Leadership

Demonstrated applied ML expertise and decision-making throughout the project lifecycle.



Key Takeaways & Next Steps

Key Area	Highlights
Model Performance	XGBoost achieved lowest MSE (211.76), $R^2 = 0.86$
Business Impact	Enables accurate, scalable weight prediction for healthcare and analytics
Actionable Insights	Supports targeted interventions, resource planning, and risk stratification
Scalability	Ready for deployment across diverse datasets and business domains
Next Steps	Integrate into workflows, monitor outcomes, expand to new use cases