**Introduction**

obesity is a significant global health concern, contributing to various chronic diseases such as cardiovascular ailments, diabetes, and certain cancers. Accurate prediction of obesity risk is crucial for early intervention and effective management. This report explores the application of machine learning models to predict obesity risk, utilizing a dataset encompassing various factors influencing obesity.

**Data Description and Preprocessing**

The dataset, sourced from Mexico, Peru, Colombia (<https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition>), comprises features related to dietary habits, physical activity, and other factors influencing obesity. Key features include demographic information (age, gender), physical measurements (height, weight), eating habits (frequency of high-caloric food intake), and physical activity levels. The target variable categorizes individuals into different obesity levels defined by the World Health Organization. Before analysis, data preprocessing steps were undertaken to ensure quality and consistency. This involved handling missing values, encoding categorical variables, and normalizing numerical features to facilitate practical model training.

**Exploratory Data Analysis**

Exploratory Data Analysis (EDA) was conducted to uncover patterns and relationships within the data. Visualizations revealed correlations between specific features and obesity levels. For instance, a higher frequency of high-calorie food intake and lower physical activity levels were associated with increased obesity risk. These insights guided the feature selection process for model development.

**Model Development and Evaluation**

our machine learning models were developed to predict obesity risk: Random Forest, XGBoost, Decision Tree, and Gradient Boosting. The performance of these models was evaluated using metrics such as accuracy, precision, recall, and F1 score.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| Random Forest | 0.988 | 0.9948 | 0.9795 | 0.9871 |
| XGBoost | 0.9976 | 1.0000 | 0.9949 | 0.9974 |
| Decision Tree | 0.9928 | 1.0000 | 0.9846 | 0.9922 |
| Gradient Boosting | 0.9952 | 0.9949 | 0.9949 | 0.9949 |
| **Conclusion**  Based on the combined evaluation of performance metrics, ROC and AUC scores, as well as the cross-validation results, XGBoost emerges as the best model for predicting the risk of obesity. It consistently demonstrated high performance, making it the most reliable and effective models among the ones evaluated. |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

.