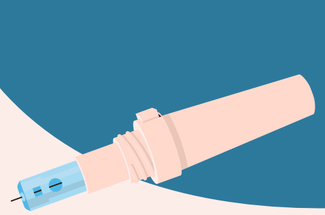
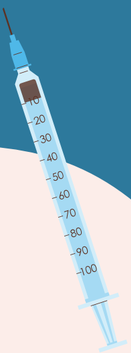
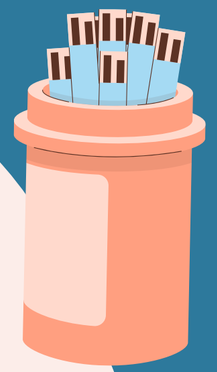
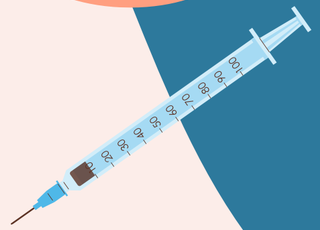
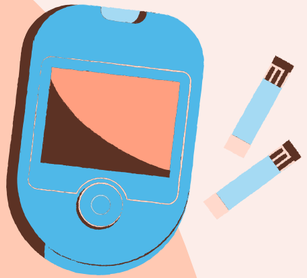


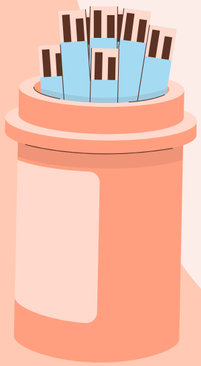


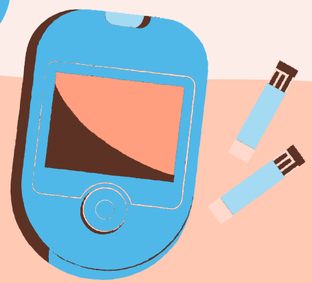




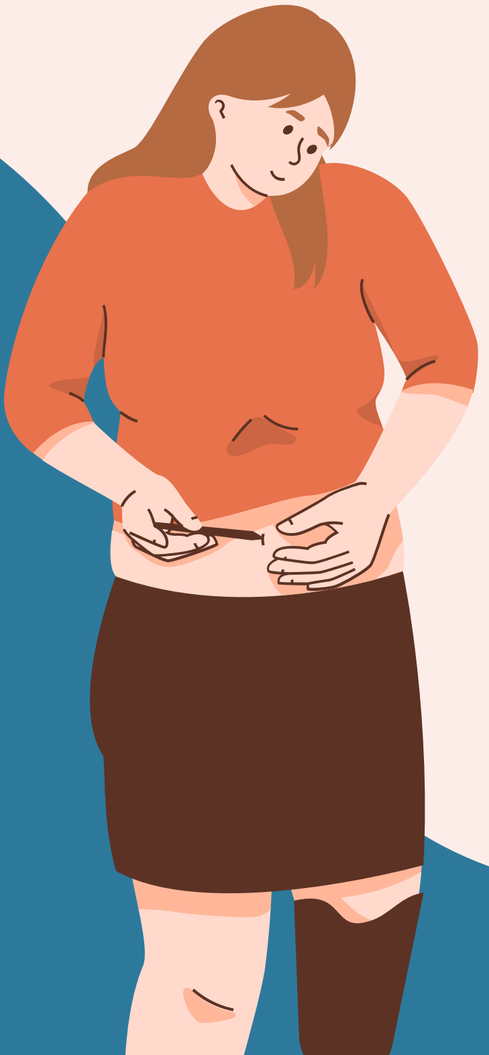
OBESITY LEVEL PREDICTION AND RECOMMENDATION SYSTEM







WHAT IS OBESITY?



Obesity refers to a state of being significantly overweight, where an individual's body fat levels are notably higher than what is considered healthy. Obese individuals are at greater risk for various health issues, including heart disease, diabetes, and certain cancers.

PROBLEM STATEMENT

"Developing an advanced system to accurately predict an individual's risk of obesity based on demographic, lifestyle, and health data, and offering personalized recommendations to prevent or manage obesity through targeted diet, exercise, and behavioral interventions."

RELEVANCE

- Obesity is the major cause of Cardiovascular diseases, Diabetes, respiratory problems and many more. It also influences quality of life such as sleeping or moving. Refer here
- This system will provide an early intervention so that individuals can properly plan their lifestyle habits and mitigate the risk of these potential problems that arises due to obesity.
- The system can raise awareness about obesity-related risks and empower individuals with knowledge to make informed decisions about their health.

APPROACH

This is a comprehensive Machine Learning model comprises of supervised and unsupervised learning solving two problems.

1. **Classification Problem**: The goal is to predict an individual's classification into one of seven specific classes related to obesity levels i.e. ‘Insufficient weight’, ’normal weight’, ‘overweight level I’, ‘overweight level II’, ‘obesity type I’, ‘obesity type II’, ‘obesity type III’.
2. **Clustering Problem:** Goal is to form clusters/groups of data on some special features and then recommend some tips based on some unique charactersitics of each group. These recommendations should be personalized and targeted.

CHOOSING THE DATA

The dataset comprises of wide variety of personal factors, eating habits and lifestyle factors that can contribute in caussing a certain type of obesity in an individual. There can be many more factors other than BMI and weight that can cause obestiy like Lack of physical activity, consumption of high caloric food, amount of water consumed etc. For info refer the article .

Credits: This dataset is taken from kaggle.
The dataset encompasses estimates of obesity levels in individuals from Mexico, Peru, and Colombia, spanning ages 14 to 61, and representing diverse dietary habits and physical conditions.

WORKFLOW OF CLASSIFICATION MODEL

BASIC ANALYSIS

checking data types, null values, duplicates, basic stats etc.

EXPLORATORY DATA ANALYSIS

Performing in depth univariate and multivariate analysis on categorical and numerical columns, checking their relations with output variable, adding feature etc.

DATA PREPROCESSING

checking outliers and handling them, splitting data into train and test encoding categorical data into numerical

MODEL TRAINING

splitting train data into train and valid sets and doing experiment on various classification models.

MODEL PERFORMANCE AND ACCURACY METRICS

A variety of classification models were used to train and test training and validation data with various experiments such as fine tuning, hyperparameter tuning , important feature extraction Train data is divided into train and validation set by k-fold validation split.

Score Metrics : score was calculated by accuracy as this is a balanced classification problem

Algorithm	Training score	Validation score
Logistic Regression	67.82%	67.76%
Random Forest Classifier	92.77%	89.43%
Gradient Boosting Classifier	89.79%	88.97%
XGBoost Classifier	92.44%	89.34%

FINALLIZING ALGORITHM

On the basis of results got by few experimentations, two models were finally selected to be trained on training and testing data with or without Important features which were extracted by feature selection.
These models were Random Forest Classifier and XG Boost classifier

Finally Random Forest Classifier was chosen as a final algorithm with important columns for two main reasons:

1. **Confusion Matrix :** confusion matrix made by RF was better than XG boost as it was less loss situation then XG boost.
2. **Time**: training time taken by random forest was much lesser than XG boost former is a bagging technique and later is boosting one.

WORKFLOW OF CLUSTERING MODEL

After encoding of data a number of experiments was performed on data to make the proper clusters and choosing the appropriate value of K

Preprocessing steps

1. On full data
2. Without output column
3. Minmax and standard scaling
4. Principal Component Analysis
5. With important Columns on which classification model is trained

ALGORITHMS USED:

1. K-Means Clustering(elbow plot and silhouette score
2. DB Scan (silhouette score and no. of clusters)
3. Hierarchical clustering (Dendogram and silhouette score)

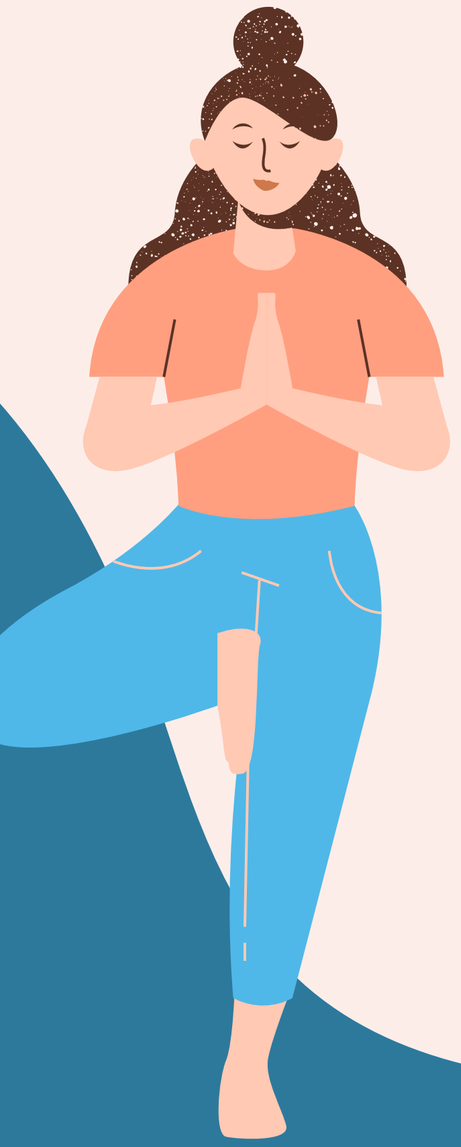
FINALLIZING ALGORITHM

On the basis of results got by few experiments K-Means clustering algorithm was finalized with 4 clusters . Model is trained on important columns same in classification model as we were getting almost same results on full data and these features as it is good to make a less complex model.

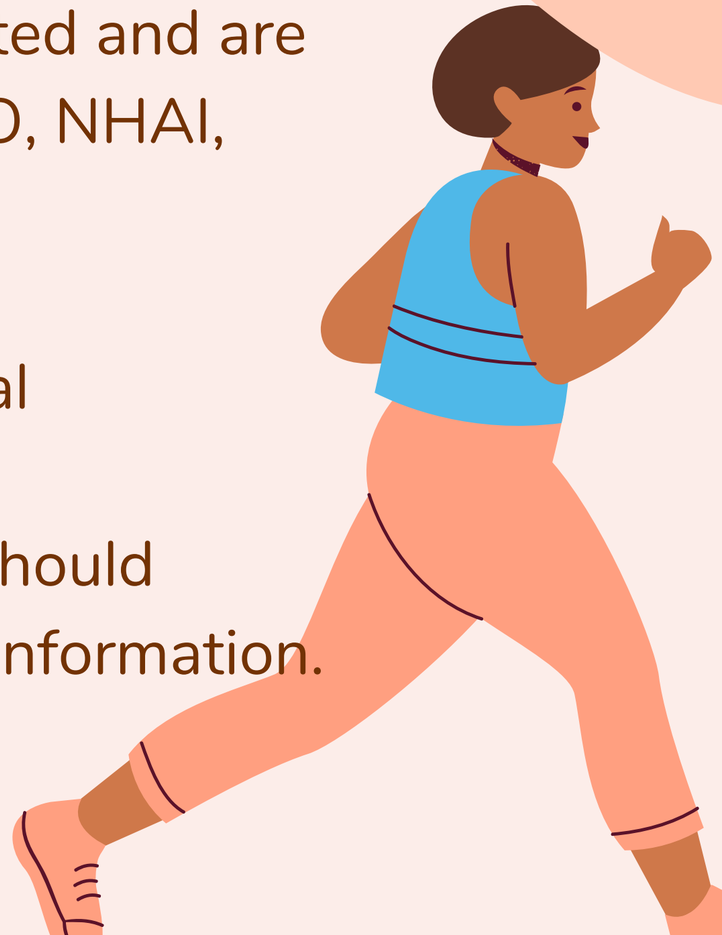
Reasons for choosing 4 clusters:

1. **Elbow Plot:** At 4 no clusters elbow plot was quite lower and slope after 4 was lesser in comparison to 3-4 and after 5 is was more lesser.
2. **Silhouette Score**: at 4 clusters score is 0.5 which is lesser than 0.62 score at 2 clusters but we cannot go for 2 clusters for such a wide data
3. **Proper pattern in clusters**: after making 4 clusters when data was divided into 4 clusters then there was a pattern seen in each cluster.

RECOMMENDATIONS



- Recommendations written on the basis of some unique characteristics of each cluster.
- These recommendations are well researched cited and are taken from some very reliable sources like WHO, NHAI, research work of Harvard University.
- These are very generic and lifestyle related recommendation and does not have any medical suggestions that can harm anyone
- Although recommendations are there but one should consult a doctor or health care expert for more information.



FRONTEND AND MODEL DEPLOYMENT

A user friendly frontend is prepared through “Streamlit”, a python library used to make frontend in which you can fill in your details and it will return your obesity level l and also give recommendations based on your information.

Information Input by user

OBESITY LEVEL

obesity level prediction with random forest classifier returns obesity level and percentage of sureity.

RECOMMENDATION SYSTEM

giving recommendation one the basis on cluster in which a individual lies.

POTENTIAL USE CASES

- Can be easily used by any user as it is very user friendly and can give a idea about his health and can protect him from possible future threat.
- Can be used in corporates to assess health of their employees as corporate culture requires a lot of sitting hours that can lead to health deprecation and further it can introduce some activities ton enhance performance and health of employees.
- can be used in schools and colleges.

THANK YOU!

Thank you so much for watching our presentation! Do you have any questions, comments, or suggestions?