OBESITY DETECTION – SUMMARY

PAPER -1

Title of the paper	Ferdowsy, Faria, Kazi Samsul Alam Rahi, Md Ismail Jabiullah, and Md Tarek Habib. "A machine learning approach for obesity risk prediction." *Current Research in Behavioral Sciences* 2 (2021): 100053.
Area of work	Prediction of Obesity
Dataset	Collected 1100 data based on 28 factors and then labelled the class of each record of the data set by consulting with some nutritionists and student counsellors in educational institutions.
Methodology / Strategy	k-NN, Support Vector Machine, Logistic Regression, Naïve Bias, Random Forest, Decision Tree, ADA Boosting, MLP, Gradient Boosting are compared on the basis of Accuracy, Sensitivity, Specificity, Precision Recall and F1-score.
Algorithm	k-NN, Support Vector Machine, Logistic Regression, Naive Bias, Random Forest, Decision Tree, ADA Boosting, MLP, Gradient Boosting.
Result/Accuracy	Logistic Regression 97.09% Naive Bayes 86.04% k-NN 77.50% Random forest 72.30% Decision tree 70.30% ADA boosting 70.03% MLP 66.02% Support Vector Machine 66.02% Gradient boosting 64.08%

PAPER – 2

Title of the paper	Rodríguez, Elias, Elen Rodríguez, Luiz Nascimento, Aneirson Francisco da Silva, and Fernando Augusto Silva Marins. "Machine learning Techniques to Predict Overweight or Obesity." In IDDM, pp. 190-204. 2021.
Area of work	Prediction of Overweight or Obesity
Dataset	The data set for this study was collected through a survey, in which 16 questions related to the interviewees' dietary habits and physical condition were applied.
Methodology / Strategy	Decision Tree, Support Vector Machines, K Nearest Neighbours, Gaussian Naive Bias, Multilayer Perceptron, Random Forest, Gradient Boosting, Extreme Gradient Boosting are compared on the basis of Accuracy, Precision, Recall and F1-score.
Algorithm	Decision Tree, Support Vector Machines, K Nearest Neighbors, Gaussian Naive Bias, Multilayer Perceptron, Random Forest, Gradient Boosting, Extreme Gradient Boosting.
Result/Accuracy	Random Forest 77.69% Gradient Boosting 73.43% Decision Tree 72.62% Extreme Gradient Boosting 70.06% K-Nearest Neighbors 67.69% Multilayer Perceptron 63.77% Support Vector Machines 59.45% Gaussian Naive Bayes 46.24%

PAPER - 3

Title of the paper	Solomon, D. D., Khan, S., Garg, S., Gupta, G., Almjally, A., Alabduallah, B. I., ... & Abdallah, A. M. A. (2023). Hybrid Majority Voting: Prediction and Classification Model for Obesity. *Diagnostics*, *13*(15), 2610.
Area of work	Obesity Prediction and Classification Model
Dataset	The dataset was taken from the UCI Machine Learning Repository. It consists of 2111 records with 17 features related to eating habits and physical conditions collected from Mexico, Peru, and Colombia.
Methodology / Strategy	k-NN, Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, Gradient Boosting, Extreme Gradient Boosting, MLP, Gaussian Naïve Bayes are compared on the basis of Accuracy, Precision Recall, F1-Score.
Algorithm	k-NN, Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, Gradient Boosting, Extreme Gradient Boosting, MLP, Gaussian Naïve Bayes.
Result/Accuracy	Support Vector Machine 86.75% Gaussian Naive Bayes 88.17% k-Nearest Neighbours 78.23% Logistic Regression 86.91% Decision Tree 94.95% Random Forest 91.95% Extreme Gradient Boosting 96.37% Gradient Boosting (GBoost) 96.06% Multilayer Perceptron (MLP) 93.38% Hybrid Model (XGBoost, GBoost, MLP) : 97.16%

PROJECT PROPOSAL

From the above three papers, we get to know that different approaches are used for obesity detection. From the observation it is found out that Logistic Regression and Random Forest are more accurate than the other algorithms with approximate accuracy of 97% and 78% respectively.