Minor-Project-Readme-File

Title of the Project : BODY FAT ESTIMATOR USING ENSEMBLE METHODS

Project Idea :

To develop a novel machine learning model that integrates SVR(Support Vector Regression), Random Forest and boosting algorithms (AdaBoost, Gradient Boosting Machine, XG Boost) for precise and cost-effective body fat percentage prediction.

Modules included in our Project:

1) Data Preprocessing
2) Feature Extraction
3) Model Training
4) Model Evaluation

Data Preprocessing:

Data preprocessing is the process of preparing and transforming raw data to make it suitable for analysis and modeling tasks.
Here, the input data are handling the missing data, normalizing, encoding numerical features, resizing and transforming into a suitable format (numpy arrays) for further model training. Additionally, data splitting help in preparing the data into separate sets for training and evaluation.

Feature Extraction:

Feature extraction is done using a pre-trained model like Extra Tree Regressor which helps in capturing relevant and useful features from the dataset. It leverages the learned representations from a large dataset, saving time and resources. These extracted features can be used as input for training the model (in this case, Gradient Boosting) to perform the regression task effectively.

Model Training:

Model training is the process of training a machine learning model on a labeled dataset to learn the underlying patterns and relationships between the input features (in this case, extracted dataset features) and the target variable.

Model is trained by using different algorithms such as Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), Ada Boost, Gradient Boosting, Artificial Neural Network (ANN).

Model Evaluation:
Model evaluation is performed to compare the performance of the algorithms (Decision Tree, Random Forest, Ada Boost, Gradient Boosting, Support Vector Regression, Artificial Neural Network) to predict the body fat percentage.The r2 score of each algorithm is calculated, compared to determine which algorithm performs the best. Gradient Boosting obtained the highest r2 score among all the algorithms. This indicates that Gradient Boosting achieved the highest percentage of correct predictions on the test set compared to the other algorithms. This result suggests that Gradient Boosting is the most effective algorithm for predicting the accurate body fat percentage prediction.

Area Identification:
Ensemble method selection plays a vital role in area detection in Body Fat Prediction. Ensemble methods can be employed for predicting body fat percentage with various input parameters like abdomen, chest, hip, weight and other body measurements. The ensemble models can be fine-tuned using techniques like Randomized Search CV, and the final model's performance can be evaluated using metrics like RMSE, R-squared, and rRMSE.. Final Visualization such as actual versus predicted plots can aid in assessing the model's performance.

Conclusion:
Using Ensemble methods for body fat estimation can provide the improved results and robustness, with potential applications in fitness, healthcare, and feature importance analysis. Careful data preprocessing and model selection are key to their success.

In Future, the prediction of body fat percentage prediction using the proposed model includes assessing the optical data-based model's performance, investigation of gender effects on BFP and testing its efficiency for predicting body fat percentage in obese children.