Human Activity Recognition Using Smartphones: Feature Engineering Techniques and Stacking/Ensemble models for accurate classification

Name:	**Priyanshu Kumar Rai**
Registration No./Roll No.:	20215
Institute/University Name:	IISER Bhopal
Program/Stream:	EECS
Problem Release date:	January 10, 2023
Date of Submission:	March 1, 2023

1 Introduction

This project aims to develop a model that accurately recognizes human activity based on data collected from smartphones. The dataset contains 563 different features and 8239 data points. In this project, we will be using different feature engineering techniques and classification models to accurately classify human activity into one of the six classes.

2 Plans to solve the problem

We plan to classify human activities using various machine learning models from the scikit-learn library, namely ***k*-Nearest Neighbors (*k*-NN)**, **Support Vector Machine (SVM)**, **Random Forest**, **Decision Tree**, and **Multilayer Perceptron (MLP)**. For accurate predictions, we will be using the method of **stacking/ensemble** of classifiers. The stack will include a selection of the best-performing classifiers from the listed models. Initially, we stacked k-NN, SVM, and Random Forest classifiers. The idea is to individually carry out hyperparameter tuning for the classifiers before stacking to get the most optimized parameters for the individual classifiers. We are keen on using the evaluation metrics of **accuracy**, **Matthew's correlation coefficient (MCC)**, **F1 score**, and **confusion matrix** for analysing the performance of each of the classifiers and the designed ensemble model.

The feature engineering techniques we plan to use for the dataset include **low-variance thresholding** [1] and **minimum redundancy maximum relevance (mRMR)** [2]. Low-variance thresholding will eliminate features below a particular variance. To determine the optimum threshold value, we passed a range of values for variance, obtained a plot for the same and finally selected the one for which we achieved maximum accuracy. The mRMR is a feature selection method that eliminates features by evaluating the mutual information to get the value of relevance and redundancy [3]. We used a range of values to identify the K best features with the highest accuracy according to the mRMR algorithm outlined in the paper. A plot of accuracy versus K best features was then generated. We plan to explore further engineering techniques to get better results.

3 Individual Contributions

The simulation for mRMR feature engineering technique and implementing Random Forest, Decision Tree and SVM and performance evaluation was done by me. The stacking/ensemble classifier model was mutually discussed upon and carried out.

References

[1] "sklearn.feature_selection.VarianceThreshold." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.VarianceThreshold.html

[2] A. Doewes, S. E. Swasono, and B. Harjito, "Feature selection on human activity recognition dataset using minimum redundancy maximum relevance," in *2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*. IEEE, jun 2017.

[3] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, aug 2005.