

Machine Learning applied to Datasets of Human Activity Recognition - An Applications of Data Analysis in HealthCare.

Ariza-Colpas. Paola Patricia * a, b , Vicario Enrico c , Butt Aziz Shariq d , De-la_Hoz-Franco, Emiroa , Piñeres-Melo, Marlon Albertoe , Oviedo-Carrascal Ana Isabelb , Muhammad Imran Tariqd , Johanna Karina García Restrepoa and Patara Fulvio

Abstract

Background: In order to remain active and productive, older adults with poor health require a combination of advanced methods of visual monitoring, optimization, pattern recognition, and learning, which provide safe and comfortable environments and serve as a tool to facilitate the work of family members and workers, both at home and in geriatric homes. Therefore, there is a need to develop technologies to provide these adults autonomy in indoor environments. **Objective:** This study aimed to generate a prediction model of daily living activities through classification techniques and selection of characteristics in order to contribute to the development in this area of knowledge, especially in the field of health. Moreover, the study aimed to accurately monitor the activities of the elderly or people with disabilities. Technological developments allow predictive analysis of daily life activities, contributing to the identification of patterns in advance in order to improve the quality of life of the elderly. **Methods:** The vanKasteren, CASAS Kyoto, and CASAS Aruba datasets were used to validate a predictive model capable of supporting the identification of activities in indoor environments. These datasets have some variation in terms of occupation and the number of daily living activities to be identified. **Results:** Twelve classifiers were implemented, among which the following stand out: Classification via Regression, OneR, Attribute Selected, J48, Random SubSpace, RandomForest, RandomCommittee, Bagging, Random Tree, JRip, LMT, and REP Tree. The classifiers that show better results when identifying daily life activities are analyzed in the light of precision and recall quality metrics. For this specific experimentation, the Classification via Regression and OneR classifiers obtain

the best results. Conclusion: The efficiency of the predictive model based on classification is concluded, showing the results of the two classifiers, i.e., Classification via Regression and OneR, with quality metrics higher than 90% even when the datasets vary in occupation and number of activities.

Keywords: HAR, smart environment, classification techniques, vanKasteren dataset, CASAS Kyoto, CASAS Aruba.