Human Activity Recognition System: Unveiling Human Movement Through Machine Learning

This project delves into the fascinating realm of human activity recognition, exploring the power of machine learning to decipher human movement based solely on smartphone sensor data. The data, a treasure trove of information generously provided by the UCI Machine Learning Repository:
https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones,
captures the silent symphony of accelerometers and gyroscopes worn by participants as they engaged in their everyday activities.

Before unleashing the power of machine learning models, a crucial step involved data preprocessing. Here, we strategically combined the training and testing sets. This decision, akin to merging puzzle pieces, allowed for a more comprehensive evaluation of the models by enriching the data pool available for training. Feature engineering, the art of crafting even more informative features from existing data, wasn't explicitly required in this instance. The dataset itself was already brimming with rich features meticulously extracted from the sensor readings. These features encompassed a spectrum of mathematical measures, including means, standard deviations, and even components derived from the frequency domain, painting a vivid picture of the underlying sensor behavior.

With the data meticulously prepared, we embarked on the exciting journey of exploring various machine learning models. Each model, a sophisticated tool in our analytical arsenal, aimed to crack the code and predict human activities based on the sensor data:

Logistic Regression: This linear model emerged as the champion, achieving a remarkable overall accuracy of 98.03%. It didn't just hit the bullseye; it consistently delivered exceptional precision, recall, and F1-scores across all six activity classes. In simpler terms, it not only excelled at correctly identifying activities but also demonstrated a remarkable ability to distinguish between them, ensuring minimal confusion.

Linear Kernel SVM: This Support Vector Machine, wielding a linear kernel as its weapon, put up a valiant fight, achieving an accuracy of 97.86%. Its performance mirrored that of Logistic Regression very closely, with stellar metrics across all activity classes. It proved to be a reliable alternative, offering exceptional classification prowess.

RBF Kernel SVM: This SVM, equipped with a more complex Radial Basis Function kernel, secured an accuracy of 97.44%. While still performing admirably, it exhibited a slightly higher number of misclassifications compared to the linear models. It functioned akin to a skilled detective, successfully solving most cases but encountering a few instances where the clues were a touch trickier to decipher.

Decision Tree: This tree-based model, with its branching decision-making process, produced an accuracy of 92.62%. Although respectable, this score fell short compared to the other models. It also displayed lower precision, recall, and F1-scores for certain activity classes. Imagine a student diligently answering most questions correctly on an exam but struggling with a few specific topics. The Decision Tree behaved similarly in this scenario.

Discussion and Conclusion:

To gauge the effectiveness of these models, we primarily relied on accuracy as our compass. In a multiclass classification task like this, where the objective is to categorize activities into distinct classes, accuracy provides a straightforward and interpretable measure of how well the models perform. It essentially reflects the overall ability of the models to correctly classify activities. Additionally, confusion matrices, akin to detailed scorecards, were employed to visualize the distribution of correct and incorrect classifications for each model, offering a deeper understanding of their strengths and weaknesses.

As the dust settled on this exploration, a fascinating pattern emerged. Logistic Regression triumphantly emerged as the most effective model for classifying human activities within this dataset. Its exceptional accuracy, coupled with consistently high precision, recall, and F1-scores for all activity classes, solidified its position as the champion. Linear Kernel SVM followed closely, presenting a compelling alternative with minimal misclassifications. The RBF Kernel SVM, while performing admirably, displayed a slight increase in errors compared to the linear models. Finally, the Decision Tree, although achieving reasonable accuracy, exhibited lower performance metrics for some activity classes.

This project serves as a testament to the remarkable potential of machine learning techniques in the realm of human activity recognition. By leveraging smartphone sensor data, we were able to unlock valuable insights into human movement. Logistic Regression, with its robust performance, proved to be a powerful tool for this task. As we move forward, future endeavors could delve into the world of feature selection techniques, meticulously choosing the most informative features to further enhance model performance. Additionally, investigating the capabilities of deep learning models, known for their prowess in handling complex data patterns, could be another captivating avenue to explore on this captivating journey of deciphering human movement through the lens of machine learning.