Case Study: Human Activity Recognition for Enhanced User Experience

Introduction: In today's digital era, smartphones and wearable devices have become ubiquitous, offering a myriad of functionalities to users. One of the emerging areas of interest is Human Activity Recognition (HAR), where machine learning models are deployed to automatically detect and classify human activities based on sensor data collected from these devices. This technology finds applications in various domains, including health monitoring, fitness tracking, context-aware computing, and user behavior analysis.

Objective: The primary objective of this case study is to develop a robust HAR system capable of accurately recognizing and classifying human activities using sensor data from smartphones. By leveraging machine learning algorithms, I aim to enhance the user experience by providing personalized services, such as activity tracking, behavior analysis, and context-aware recommendations.

Dataset Selection: For this study, I have chosen the Human Activity Recognition (HAR) dataset obtained from the UCI Machine Learning Repository

Dataset Overview:

- The HAR dataset contains sensor data collected from the accelerometer and gyroscope sensors of smartphones worn by subjects while performing various activities.
- The activities include walking, walking upstairs, walking downstairs, sitting, standing, and laying.
- The sensor data is represented as time-series signals, where each signal corresponds to a specific feature captured by the sensors (e.g., acceleration along different axes).

Dataset Structure: The dataset is organized into the following files and directories:

1. **README.txt:**
    - Provides information about the dataset, including the data collection process, feature descriptions, and activity labels.
2. **activity_labels.txt:**
    - Contains a mapping of numerical activity labels to descriptive activity names (e.g., 1 corresponds to "walking").
3. **features_info.txt:**
    - Describes the features extracted from the sensor data and their corresponding measurements.
4. **train/ and test/ Directories:**
    - Contain separate training and testing datasets.
    - Each directory includes the following files:
    - **subject_train.txt / subject_test.txt:**
        - Contains the subject IDs corresponding to each sample in the training/testing data.
    - **X_train.txt / X_test.txt:**
        - Contains the feature vectors representing the sensor data for each sample.
    - **y_train.txt / y_test.txt:**
        - Contains the activity labels corresponding to each sample in the training/testing data.
5. **Inertial Signals Directories:**
    - Contain raw sensor signals (e.g., acceleration and angular velocity) collected during data recording.
    - These directories are optional and may not be used in all analyses.

Data Preprocessing:

- We preprocessed the dataset to ensure consistency and compatibility for model training.
- Standardization: We standardized the features using the StandardScaler from scikit-learn to scale each feature to have a mean of 0 and a standard deviation of 1.
- One-Hot Encoding: We used one-hot encoding to convert categorical activity labels into binary vectors, facilitating model training.

Privacy-Preserving Techniques:

- I employed differential privacy mechanisms to ensure that sensitive user data remains protected during model training.
- Federated learning techniques enabled collaborative model training across distributed devices while preserving data privacy.

Model Development:

- I developed a deep learning model using TensorFlow/Keras to classify human activities based on sensor data.
- The model architecture consisted of multiple layers of densely connected neurons, optimized using privacy-preserving differential privacy SGD (DP-SGD) algorithms.

Evaluation Metrics:

I evaluated the model's performance using various techniques, including confusion matrices, and classification reports.

Results: Our model achieved an accuracy of 83.64% on the testing data, indicating its effectiveness in classifying human activities. The confusion matrix revealed the model's performance across different activity classes, while the classification report provided detailed metrics for each class.

Conclusion: In conclusion, my project successfully developed a machine-learning model for human activity recognition using smartphone data. The model demonstrated high accuracy and robust performance across multiple evaluation metrics. Future work may involve exploring additional feature engineering techniques, optimizing hyperparameters, and deploying the model in real-world applications.

Potential use-cases

Integration of Human Activity Recognition technology into smartphones enables seamless tracking of user's daily activities, empowering personalized health insights and enhancing overall user experience.