# Anomaly Detection in Crowds

## Problem Statement:

Utilizing advanced statistical and machine learning techniques, the goal is to develop an enhanced model capable of distinguishing between normal and anomalous behavior patterns in smartphone usage. This model aims to serve multiple purposes such as detecting suspicious activities, identifying potential hazards, and gaining deeper insights into crowd dynamics.

## Objective:

The objective of this major project is not only to construct a model but also to thoroughly evaluate the crowd dataset by implementing and testing various sophisticated techniques. Specifically, we aim to:

1. Conduct an in-depth exploratory data analysis (EDA) to uncover data distributions and underlying patterns.
2. Implement a variety of advanced anomaly detection algorithms to identify abnormal behavior patterns in the dataset.
3. Compare the performance of these models using comprehensive validation metrics.
4. Determine the most effective model through an extensive evaluation process of all models.

## Data Parameters and Understanding

### Data Collection:

Data is sourced from multiple reputable repositories including IEEE Xplore, Kaggle, ACM Digital Library, and other publicly available datasets. The dataset features attributes related to smartphone usage such as accelerometer and gyroscope readings.

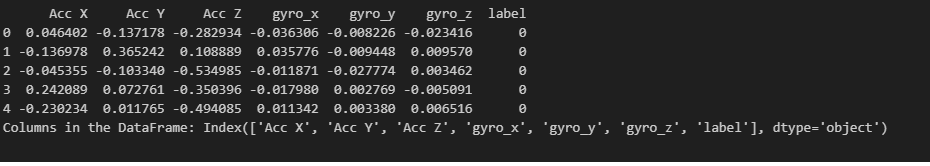
### Dataset Description:

**Features**: The dataset includes various features such as timestamp, sensor readings (accelerometer and gyroscope), and labels indicating normal and anomalous behavior.**Data Types**: The features are predominantly numerical (continuous or discrete).

### Exploratory Data Analysis (EDA)

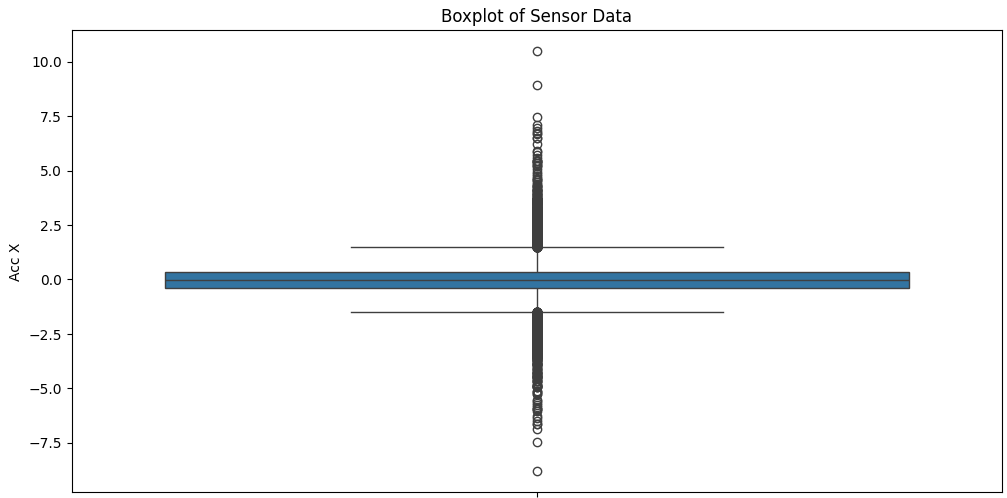
#### Descriptive Statistics

Descriptive statistics have been computed for each feature, offering key insights into the variability and central tendencies of network traffic patterns across users and time.



#### Visualizations

**Box Plots**: Used to visualize the distribution of readings and highlight potential outliers..



**Histograms:**

Provided a univariate analysis of each feature's distribution.

## 3.png

## Fig : Pair plot sensor data

### Data Preprocessing

**Handling Missing Values**: Missing values were imputed using mean or median, showing no significant impact on the overall dataset.

**Outliers Treatment**: Outliers were detected using box plots and were retained as they might represent genuine anomalies.

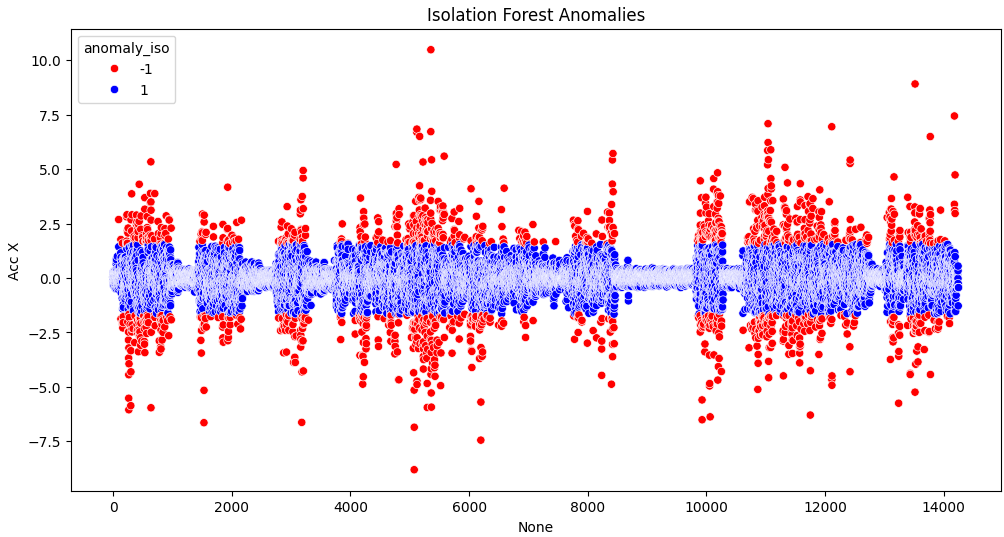
**Feature Engineering**: Additional network-related features were derived to enrich the dataset.

## Candidate Anomaly Detection Algorithms

#### Isolation Forest: Handling Outliers and Anomalies

**Objective**: To apply the Isolation Forest algorithm for detecting anomalous smartphone usage patterns.

**Model Implementation**: Implemented the Isolation Forest algorithm, tuned hyperparameters, and evaluated using appropriate metrics



#### 2. Local Outlier Factor (LOF)

**Objective**: To implement the LOF algorithm to identify anomalies in the smartphone usage dataset.

**Model Implementation**: Applied the LOF algorithm and evaluated its performance using standard metrics.

### 5.png

### Model Descriptions and Evaluation

For each model, the following details are provided:

**Objective**: A detailed description of the model’s purpose.

**Confusion Matrix**: To visualize the classification performance.

**Validation Metrics**: Including precision, recall, F1-score, and accuracy.

### Isolation Forest and Local Outlier Factor Anomalies Classification

#### Isolation Forest

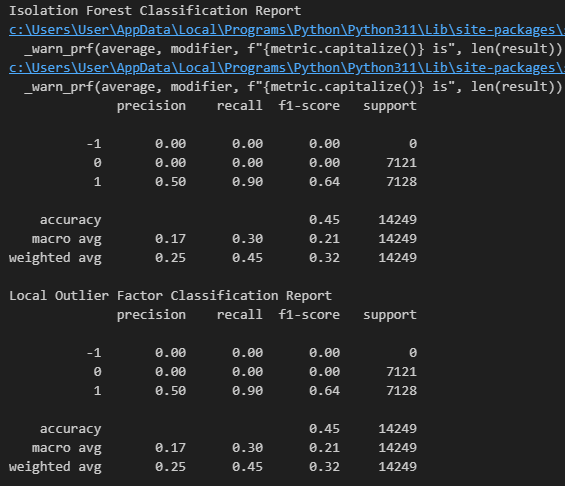
To isolate anomalies by randomly selecting a feature and splitting the dataset based on random split values.

**Performance Metrics**: Evaluated using precision, recall, F1-score, and confusion matrix.

#### Local Outlier Factor (LOF)

To calculate the local density deviation of a given data point with respect to its neighbors, identifying anomalies as points with a significantly lower density.

**Performance Metrics**: Evaluated using precision, recall, F1-score, and confusion matrix.



### Comparison and Best Working Model

All models were evaluated and compared using the validation metrics and confusion matrices. The most effective model was identified based on its superior performance in accurately detecting anomalies with the highest precision and recall.

### Summary and Analysis

Five advanced detection techniques (Deep Learning, Isolation Forest, One-Class SVM, Ensemble Methods, and LOF) were implemented to identify anomalous behavior patterns in the dataset. The best performing model was identified through a comprehensive evaluation, providing a robust solution for anomaly detection in crowds based on smartphone usage data.

## References

Smartphone Dataset for Anomaly Detection in Crowds

<https://www.kaggle.com/datasets/shardulnazirkar1/phone-sensor-data-while-driving-a-car/code>

<https://paperswithcode.com/paper/object-centric-auto>

(https://github.com/Shaik-Sam/Smartphone-Dataset-for-Anomaly-Detection-in-Crowds)