﻿**DSAA 5002 - Data Mining and Knowledge Discovery in Data Science**

**Final Exam Report - Q1 ﻿Supervised Outlier Detection**

**50015940 Jiaxiang Gao**

**Main Experimental Steps and Methods:**

1. **Data Preparation:**

**Feature and Target Separation:** The feature variables (X) were separated from the target variable (Is\_Falling) in the training dataset.

**Feature Standardization:** The StandardScaler was utilized to scale the features, ensuring a uniform evaluation.

**Handling Class Imbalance:** SMOTE (Synthetic Minority Over-sampling Technique) was applied to generate synthetic samples for the minority class, thus balancing the dataset.

1. **Test Data Preparation:**

**Test Data Processing:** The test dataset was processed by dropping the 'ID' and 'Is\_Falling' columns. The 'Is\_Falling' column is the target variable to be predicted, and 'ID' is not a feature. The same scaler used for the training data was applied to scale the test data.

1. **Model Training and Evaluation:**

**3.1 Random Forest Classifier:**

**Training:** A RandomForestClassifier was trained using the resampled training data.

**Evaluation:** The classifier's performance was evaluated on the test data, focusing on metrics like precision, recall, and F1-score.

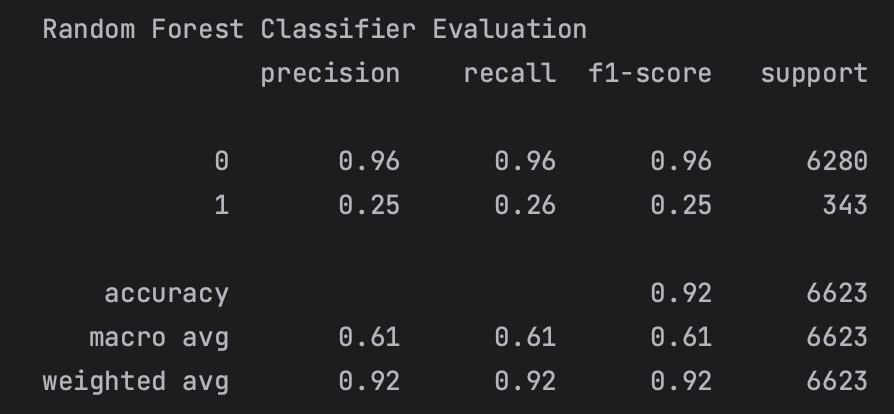
**3.2 LightGBM Classifier:**

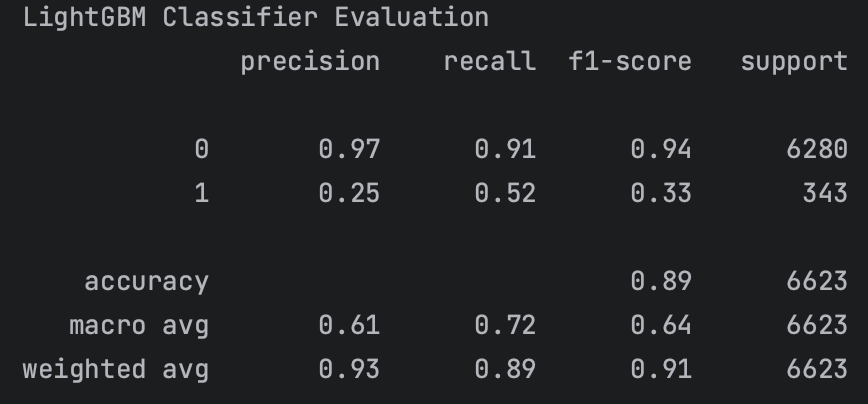
**Training:** An LGBMClassifier was similarly trained on the resampled training data.

**Evaluation:** This classifier's performance was also assessed using the same metrics as the RandomForestClassifier.

1. **Comparison and Decision:**

**Performance Analysis:** The recall rate for the minority class (falling class) was a key metric in the comparison. LightGBM demonstrated a significantly higher recall rate compared to the Random Forest, indicating its superior ability in identifying true falling events.





**Final Selection:** Based on the higher recall rate for the minority class, LightGBM was chosen for predicting the test data, especially considering the critical nature of the application (fall detection).