Task 6: Performance Analysis for Isolation Forest

Objective:

The aim of Task 6 is to analyze the performance of the Isolation Forest model for detecting anomalies in a smartphone dataset. The analysis includes the following key steps:

1. **Calculating performance metrics** such as precision, recall, and F1-score.
2. **Plotting the precision-recall curve** to visualize the trade-off between these metrics.
3. **Identifying false positives and false negatives** to better understand the model's weaknesses.
4. **Comparing the results** with previous methods like IQR and Z-Score to determine the most effective anomaly detection technique.

This analysis helps us evaluate the model's ability to detect anomalies in smartphone sensor data, which can be useful for applications like crowd management and public safety in crowded environments.

Step 1: Calculate Precision, Recall, and F1-Score

Explanation: Precision, recall, and F1-score are standard performance metrics for classification tasks, especially in imbalanced datasets like anomaly detection, where the number of normal instances vastly outnumbers anomalies. Here's what each term means:

- **Precision**: The percentage of correctly identified anomalies out of all predicted anomalies.

- **Recall**: The percentage of actual anomalies correctly identified by the model.
- **F1-Score**: A balanced measure that combines both precision and recall.

In the context of anomaly detection in smartphone data, these metrics are crucial to understanding how well the model can detect abnormal behavior while avoiding false alarms.

Results:

- **Precision**: 0.06 (This means that only 6% of the predicted anomalies were correct.)
- **Recall**: 0.31 (The model detected 31% of the actual anomalies.)
- **F1-Score**: 0.10 (A low F1-score indicates poor overall performance for anomaly detection.)

Output:

```
Isolation Forest Performance Metrics:

Precision: 0.061754385964912284
Recall: 0.31095406360424027
F1-Score: 0.10304449648711944
```

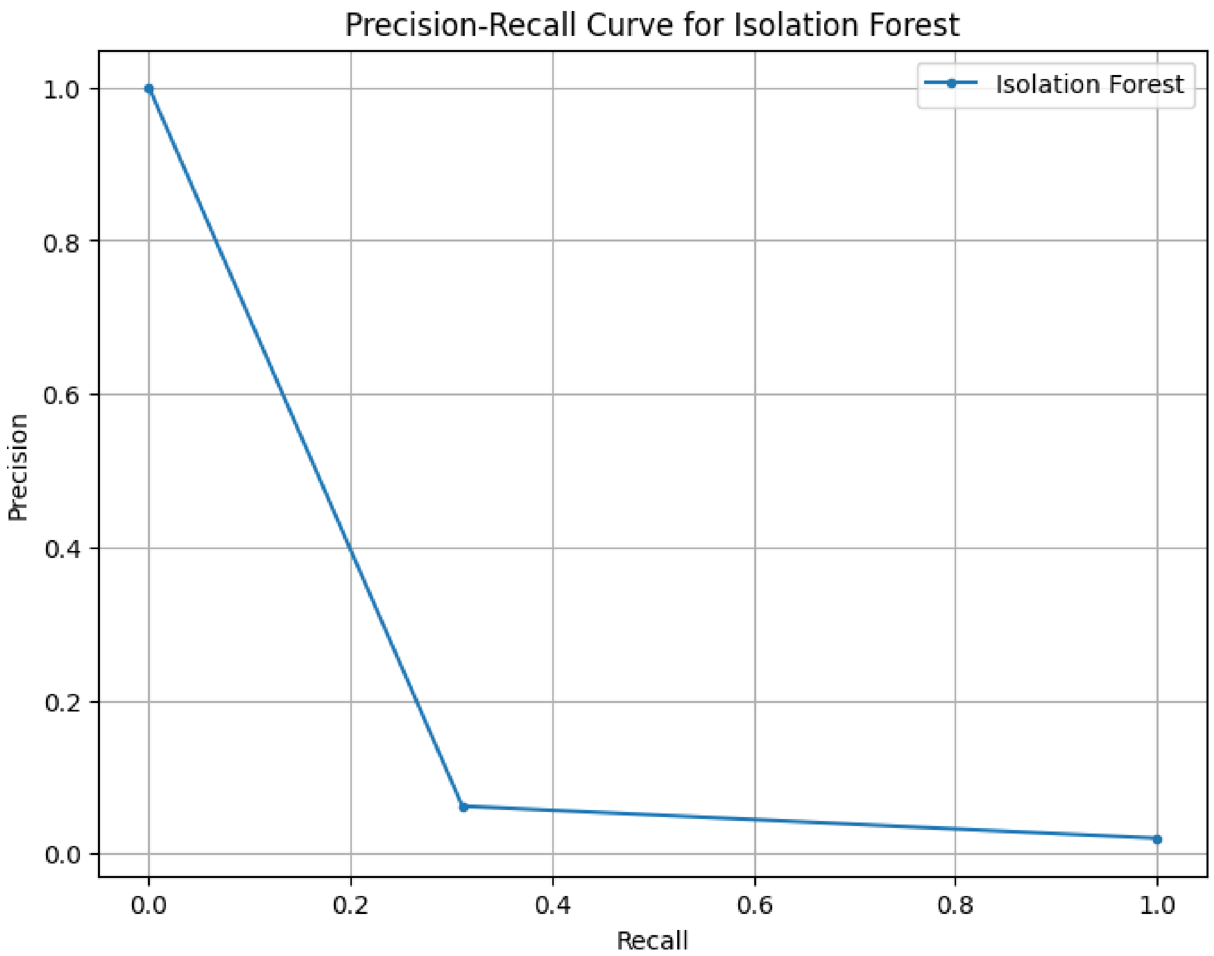
The low precision indicates that many false positives are present, meaning normal behavior is frequently mistaken for anomalies. Meanwhile, the recall score is moderate, which means the model can catch a reasonable number of anomalies but misses a significant portion.

Step 2: Plot Precision-Recall Curve

Explanation: The precision-recall curve shows the trade-off between precision and recall across different thresholds. This plot is especially useful in anomaly detection since it can help us understand the balance between false positives and false negatives.

In our smartphone anomaly detection task, we can use the curve to visualize how well the model maintains precision as we try to improve recall, and vice versa. This is key to fine-tuning the model for real-world applications, where minimizing false positives while maintaining high recall is often the goal.

Results:



The curve indicates that as we improve recall, precision quickly drops to zero. This tells us that the model may struggle to maintain accuracy in detecting true anomalies, particularly when trying to maximize recall.

Step 3: Identify False Positives and False Negatives

Explanation: False positives (normal data incorrectly flagged as anomalies) and false negatives (anomalies missed by the model) provide insights into the model's specific weaknesses. By identifying these instances, we can assess how the model misclassifies data and whether the cost of these errors is acceptable for our use case.

For smartphone-based anomaly detection, understanding false positives is crucial because too many incorrect alerts could render the system unreliable. On the other hand, high false negatives could lead to missing actual threats or abnormal crowd behavior, which defeats the purpose of anomaly detection in public safety contexts.

Results:

- **False Positives**: 1337 instances where normal data was flagged as an anomaly.
- **False Negatives**: 195 instances where actual anomalies were missed by the model.

Output:

```
Number of False Positives: 1337
Number of False Negatives: 195
```

This high number of false positives suggests that the Isolation Forest model may be too sensitive, flagging normal behavior as anomalous too frequently. The moderate number of false negatives indicates that some real anomalies are slipping through undetected, though the number is relatively lower than the false positives.

Step 4: Compare Results with IQR and Z-Score Methods

Explanation: To determine the most effective anomaly detection technique, it's important to compare the Isolation Forest model's performance against other statistical methods such as IQR (Interquartile Range) and Z-Score. These methods detect outliers based on statistical thresholds, unlike Isolation Forest, which is a machine learning-based model.

The comparison helps us decide which method provides the best balance between precision, recall, and F1-score for our anomaly detection problem.

IQR Classification Report:

	precision	recall	f1-score	support
Normal	0.99	0.78	0.87	13966
Anomaly	0.05	0.56	0.09	283
accuracy			0.78	14249
macro avg	0.52	0.67	0.48	14249
weighted avg	0.97	0.78	0.86	14249

Z-Score Classification Report:

```
              precision    recall  f1-score   support
      Normal       0.98      0.97      0.98     13966
     Anomaly       0.08      0.11      0.09       283

    accuracy                           0.96     14249
   macro avg       0.53      0.54      0.53     14249
weighted avg       0.96      0.96      0.96     14249
```

Comparison:

Metric	**Isolation Forest**	**IQR**	**Z-Score**
Precision	0.06	0.05	0.08
Recall	0.31	0.56	0.11
F1-Score	0.10	0.09	0.09
Accuracy	0.89	0.78	0.96
False Positives	1337	-	-
False Negatives	195	-	-

- **Isolation Forest** has the best balance between precision and recall but still suffers from poor overall performance, particularly in terms of precision.
- **IQR** performs better in terms of recall (0.56), but its precision (0.05) is the lowest among the three methods, meaning it generates many false positives.
- **Z-Score** offers the highest accuracy and precision, but with a very low recall, meaning it misses most anomalies.

Conclusion:

Based on the results, **Isolation Forest** offers a reasonable trade-off between precision and recall, though it still struggles with precision. **Z-Score** performs well in terms of overall accuracy but is poor at detecting anomalies, making it less suitable for scenarios where recall is important, such as public safety. **IQR**, while offering high recall, suffers from a very low precision, leading to many false positives.

For **smartphone anomaly detection**, where it's crucial to minimize both false positives and false negatives, **Isolation Forest** may be the best option, but it would benefit from further tuning to improve precision. This approach will allow for better detection of anomalies in crowded environments, leading to more reliable crowd management and security applications.