4/11/2024 - 11/11/2024

**WEEK 1**

We started our project by reviewing the paper titled “Two-Stage Approach With Combination of Outlier Detection Method and Deep Learning Enhances Automatic Epileptic Seizure Detection”. The paper investigates outlier detection using data from 83 epilepsy patients. We also explored the outlier detection methods used in this paper with the goal of improving them.

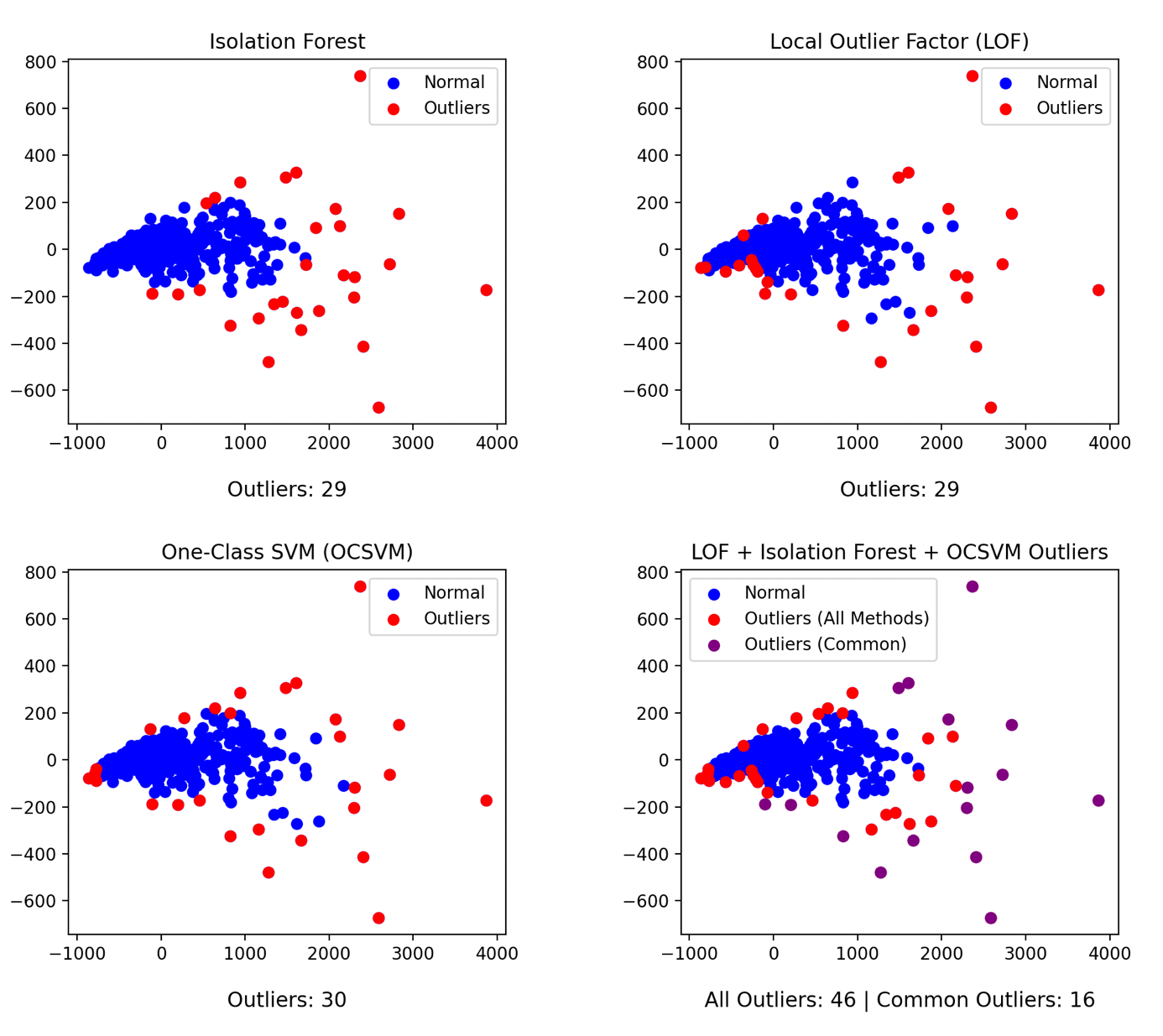
Initially, we examined the One-Class Support Vector Machines (OCSVM) method. We learned that OCSVM has been more widely used and yields more positive results when applied to healthcare-related datasets. This led us to explore healthcare datasets further. We discovered that outlier detection methods produce more accurate and precise results when applied to labeled datasets.

As a result, we decided to use the breast\_cancer dataset, which is preloaded in the sklearn library. We applied OCSVM to this dataset to obtain our first output for outlier detection.

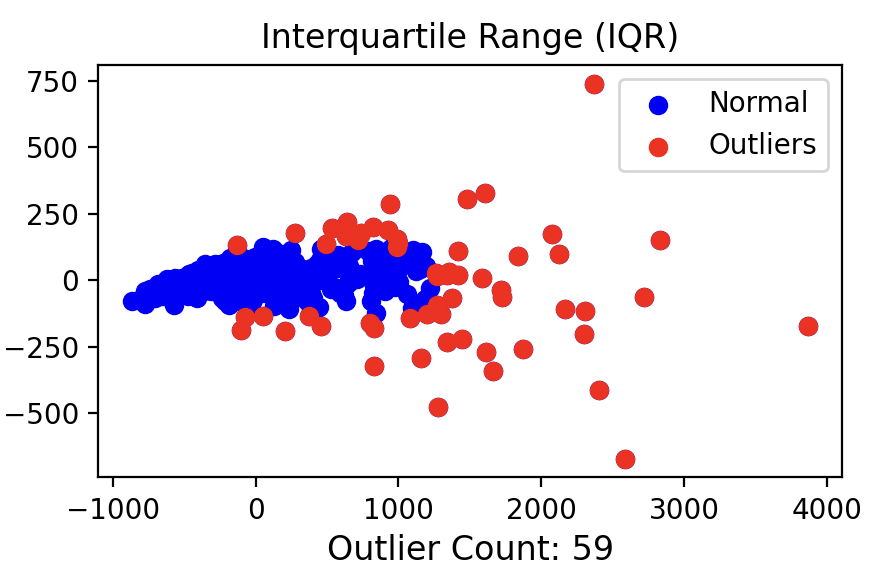
metin, ekran görüntüsü, diyagram, yazı tipi içeren bir resim

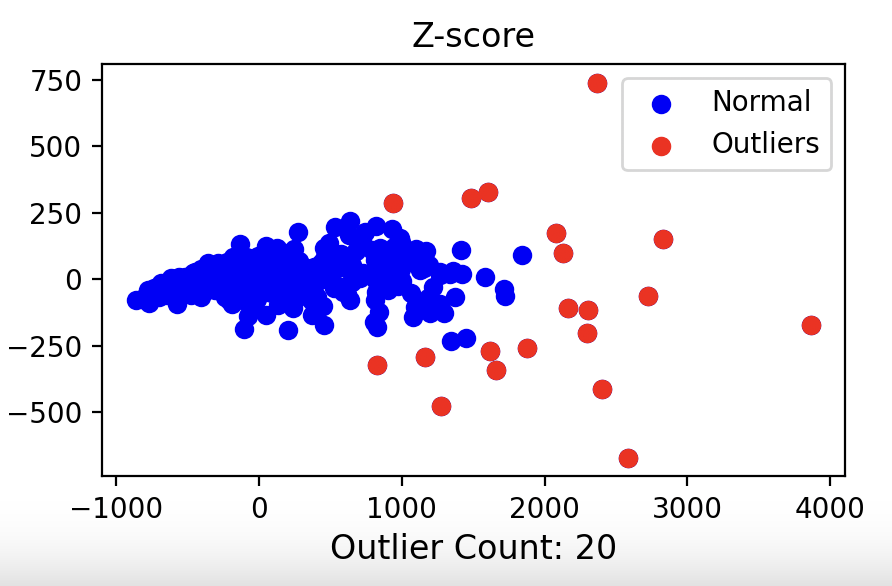
Açıklama otomatik olarak oluşturuldu

Initially, we decided to combine the OCSVM method with other methods to improve its performance. We researched other methods and applied the Isolation Forest and Local Outlier Factor (LOF) methods, which are machine learning-based techniques within the same category as OCSVM, alongside OCSVM. After applying the methods individually, we also examined the results by using all three methods together. Although the Isolation Forest and LOF methods detected fewer outliers compared to OCSVM, when we used all three methods together, we identified a larger number of outliers.



Next, we decided to combine the methods with different types of outlier detection techniques and try again. As a result of our research, we applied the Interquartile Range (IQR) and Z-Score methods to the dataset.





We added the IQR and Z-Score methods to our main code and generated a total of 6 graphs showing the outliers detected by each method individually. We observed an increase in the number of outliers, which was visible in the graph. Afterward, we continued our research to improve these methods with a different approach. We learned that the normalization or standardization of the dataset is crucial in outlier detection. To make the algorithms more efficient, we standardized the data using StandardScaler. We also made the data more compact by reducing the feature dimensions with Principal Component Analysis (PCA). Additionally, we improved the model’s accuracy and efficiency by reducing the number of features using Recursive Feature Elimination (RFE).

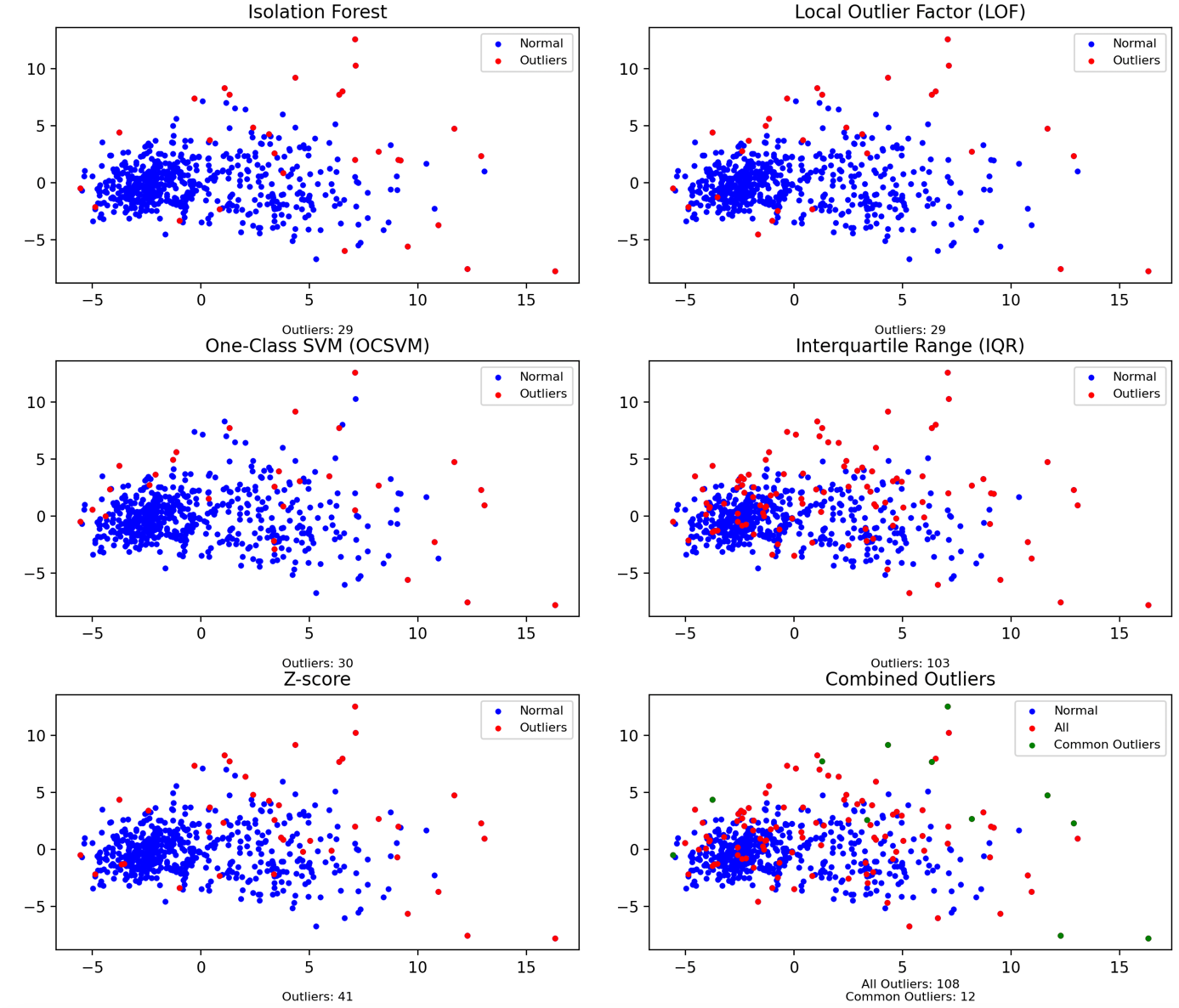
To evaluate how accurate the detected outliers were, we included Precision, Recall, and F1-Score metrics.

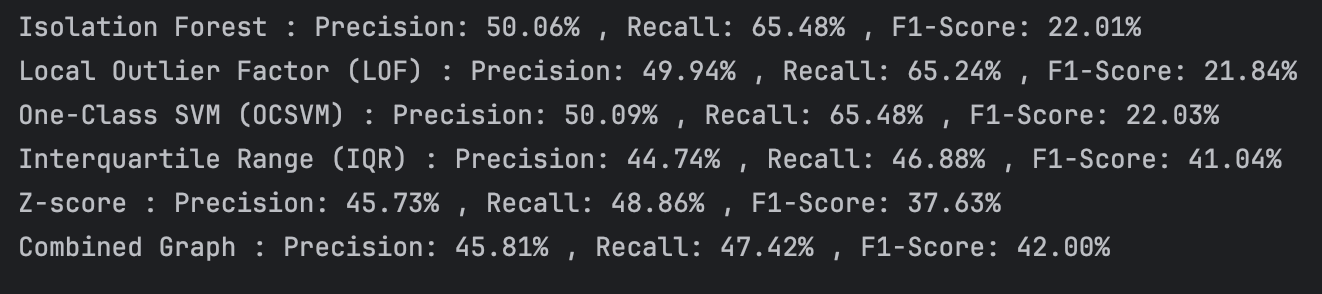
• **Precision**: Shows how many of the values flagged as “outliers” are truly outliers.

• **Recall**: Shows how many of the actual outliers were detected.

• **F1-Score**: This metric combines both precision (correct detection) and recall (ability to detect outliers).

Finally, to make the graphs cleaner and easier to visualize, we adjusted the graph sizes, font sizes, and point sizes. We obtained the final outputs and accuracy rates.





**Referencess:**

<https://scikit-learn.org/1.5/modules/generated/sklearn.datasets.load_breast_cancer.html#sklearn.datasets.load_breast_cancer>

<https://www.geeksforgeeks.org/detect-and-remove-the-outliers-using-python/>

<https://medium.com/@alex.eslava96/outlier-detection-techniques-for-time-series-9868db2875c2>

<https://www.kaggle.com/code/alexeslava/outlier-detection-techniques-for-time-series>

<https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html>

<https://medium.com/@roshmitadey/how-to-scale-data-with-a-lot-of-outliers-34c7ae52bc27>

<https://stackoverflow.com/questions/28558540/precision-and-recall-for-outlier-detection>

<https://hex.tech/templates/data-science/outlier-detection/>

<https://www.machinelearningplus.com/machine-learning/how-to-detect-outliers-with-z-score/>

<https://online.stat.psu.edu/stat200/lesson/3/3.2>

<https://www.analyticsvidhya.com/blog/2021/07/anomaly-detection-using-isolation-forest-a-complete-guide/>

<https://scikit-learn.org/1.5/auto_examples/neighbors/plot_lof_outlier_detection.html>

<https://matplotlib.org/stable/users/explain/text/text_props.html>