18/11/2024-24/11/2024

WEEK 2

We started researching methods to enhance the OCSVM method for the second week.  
As a result of our research, we discovered the Elliptic Envelope method and integrated it into our code.

ekran görüntüsü, metin, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

We continued our research and subsequently discovered the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method.  
We integrated this method into our code as well and obtained its output.

ekran görüntüsü, metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

In the 4th graph, we labeled the outliers identified by all three methods as **Combined Outliers** and those commonly identified by all methods as **Common Outliers**.

We calculated the F1 score for this graph and displayed these values below it.

The points were divided into three colors: **Combined Outliers**, **Common Outliers**, and **All Data Points**, creating the final visualization.

In total, we created 4 graphs:

1. **Graph 1**: OCSVM results
2. **Graph 2**: DBSCAN results
3. **Graph 3**: Elliptic Envelope results
4. **Graph 4**: A combined graph showing the common outliers identified by all methods.

Afterward, we continued our work to increase the number of detected outliers and improve the F1 score.

ENSEMBLE METHOD

Next, we experimented with ensemble methods by combining multiple OCSVM models to create a stronger detection approach. We trained several OCSVM models and merged their results. However, we observed a low F1 score and a decrease in the number of detected outliers. Due to these unfavorable results, we decided not to use this method in our project and continued our research.

ekran görüntüsü, metin, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

We decided to use meta-model training, another ensemble method. By using the output of each model as a feature vector, we trained a logistic regression model. This approach addressed the shortcomings of individual methods, resulting in stronger and more reliable outlier detection. For the output, we set the 4th graph to represent the meta-model. We observed that the number of outliers detected by the meta-model (29) was slightly lower than OCSVM (30) and DBSCAN (40). However, the meta-model showed a significant improvement in accuracy compared to the other methods (META MODEL: 50.78%, OCSVM: 25.82%, DBSCAN: 16.37%, Elliptic Envelope: 26.38%).

metin, diyagram, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Referencess:

<https://scikit-learn.org/dev/modules/generated/sklearn.covariance.EllipticEnvelope.html>

<https://medium.com/@dilip.voleti/dbscan-algorithm-for-fraud-detection-outlier-detection-in-a-data-set-60a10ad06ea8>

<https://scikit--learn-org.translate.goog/1.5/modules/outlier_detection.html?_x_tr_sl=en&_x_tr_tl=tr&_x_tr_hl=tr&_x_tr_pto=tc>

<https://www.datacamp.com/tutorial/understanding-logistic-regression-python?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720821&utm_adgroupid=157156375191&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adpostion=&utm_creative=720362650048&utm_targetid=aud-1832882613722:dsa-2218886984100&utm_loc_interest_ms=&utm_loc_physical_ms=9198982&utm_content=&utm_campaign=230119_1-sea~dsa~tofu_2-b2c_3-row-p1_4-prc_5-na_6-na_7-le_8-pdsh-go_9-nb-e_10-na_11-na-bfcm24&gad_source=1&gclid=Cj0KCQiAuou6BhDhARIsAIfgrn6tsAuiWdVhsujqMYCqF_scKf4SK6LSBa4Mo1JfjWCXu4Uf3-ov-C8aArVEEALw_wcB>

<https://www.geeksforgeeks.org/outlier-detection-in-logistic-regression/>