Izveštaj

Pokušao sam da primenim algoritme nenadgledanog učenja, ali nisu dali dobre rezultate.

Pretpostavljam da je zbog toga što sama priroda podataka nije pogodna za klasterovanje.

Pogledao sam odredjen broj istanci i pretpostavljam da je razlog loših rezultata ustvari sto podaci sami po sebi nisu anomalije. Podaci koji spadaju u klasu fraud upadaju u određene klastere.

Metrika koju sam primenio za evaluaciju je silhouette score koji je dao dobre rezultate što se tiče anomalija, ali loše rezultate pronalaženja pravih krađi. Zbog toga je nadgledano učenje dalo dobre rezultate.

Korišćeni algoritmi za klasterovanje su: DBSCAN, Iforest,OneClassSVM

Rezultati klasterovanja za skup card\_transdata

|  |  |  |  |
| --- | --- | --- | --- |
| Metrika | DBSCAN | Iforest | OC-SVM |
| Precision | 0.4285714 | 0.285714 | 0.167167 |
| Recall | 0.0035336 | 0.022075 | 0.972093 |
| silhouette score | 0.8853701 | 0.610817 | - |

Rezultati klasterovanja za skup hai

|  |  |  |  |
| --- | --- | --- | --- |
| Metrika | DBSCAN | Iforest | OC-SVM |
| Precision | 0.009915 | 0 | 0.015596 |
| Recall | 1 | 0 | 0.745223 |
| silhouette score | -0.35325 | 0.426176 | - |

Razlog iz kog nisam primenio algoritme nadgledanog učenja je to što ima premalo istanci koje predstavljaju napad.