Literature review

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Paper | objective | Method | Results | Advantage | Disadvantage | Different from our study |
| **Detection of Anomalous Insiders in Collaborative**  **Environments via Relational Analysis of Access Logs** | To detect insider threats based on information recorded in the access logs of collaborative environments | Community based anomaly detection system- users tend to form community structures, us users whom their behaviour are not inline with that of the community members, are indication of anomalous behaviour. The model has relational pattern extraction and anomaly detection components | CADS was able to differentiate anomalous users in a real data logs |  |  |  |
| **Mining Deviations from Patient Care Pathways via Electronic Medical**  **Record System Audits** |  | we hypothesize that care pathways can be represented as the progression of a  patient through a system and introduce a strategy to model the patient’s flow as a sequence of accesses defined  over a graph. Elements in the sequence correspond to features associated with the access transaction  (e.g., reason for access). we model patterns of patient record usage, which may  indicate deviations from care workflows. | building a graph-based model that represe nts  the series of EMR accesses resulting from intended activities related to patient  care. this framework finds a small portion of accesses  constitute outliers from such flows. also observe that the violation patterns deviate for different types of  medical services. Analysis of our results suggests greater deviation from normal access patterns by nonclinical  users. the area under the receiver operating characteristic (ROC) curve for the Pediatrics service was found  to be 0.9166. The results suggest that our approach is competitive with, and often better than, the existing  state-of-the-art in its outlier detection performance. At the same time, our method is more efficient, by  orders of magnitude, than previous approaches, allowing for detection of thousands of accesses in seconds. |  |  |  |
| **A Hybrid Density-Based Outlier Detection Model for Privacy in Electronic Patient**  **Record Systems** |  | **This paper presents a system that employs a Human-in-the-**  **Loop Machine Learning (HILML) algorithm, in addition to a**  **density-based local outlier detection model.** the LOF anomaly scores need to calculate the  ensemble average. In order to achieve this, a weighted  average is applied to each audit log. An additional column is  added next to each of the IDs with that IDs associated  anomaly score. For every audit log, a weighted average of  the four anomaly ID scores is calculated. The calculated  ensemble average anomaly score can then be plotted against  the Date & Time stamp and visualised to the analyst. | **The system is able**  **to detect 145 anomalous behaviours in an unlabelled dataset of**  **1,007,727 audit logs. This equates to 0.014% of the thEPR**  **accesses being labelled as anomalous in a specialist Liverpool**  **(UK) hospital.** ensemble  averaging and a human-in-the-loop model |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

In general, anomalies can be defined as any observations that are different from the normal behavior of the data[7]

7. Das, K., J. Schneider, and D.B. Neill. *Anomaly pattern detection in categorical datasets*. in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2008.