

Zusammenfassung

Diese Seminararbeit beschäftigt sich damit, wie mit Hilfe der beiden Lernalgorithmen „K-Means Clustering“ und „One Class SVM“ Anomalien in Zeitreihen erkannt werden können. Durchgeführt werden diese Verfahren an einem simulierten Datensatz, welcher die Besucher einer Webseite pro Tag abbilden soll. In den originalen Daten wurden die Anomalien bereits markiert. Diese Information wird lediglich bei der Auswertung und Interpretation der Ergebnisse mit dem F1-Score verwendet und kommt nicht beim Training der Lernalgorithmen zum Einsatz.

Zur Interpretation der Datenpunkte mit den Algorithmen wird die Zeitreihe zunächst in Test- und Trainingsdaten unterteilt und mit dem „Windowing“-Verfahren in kleinere Teilsegmente zerlegt. Anschließend werden diese Teilsegmente durch die beiden Algorithmen interpretiert. Nach der Interpretation werden die Segmente wieder zurück in ihre Ursprungsform rücktransformiert und mit den original gelabelten Daten abgeglichen. Als Maß der Genauigkeit wird hier der F1-Score verwendet

Abschließend wird ein Fazit zu den beiden Lernalgorithmen, in Kombination mit dem Windowing-Verfahren gezogen und Optimierungsvorschläge für zukünftige Verbesserungen gegeben.