SUMMARY

This thesis explores anomaly detection association rule mining and supervised machine learning focusing on their application to autonomous vehicle sensor data for attack detection. The objective is to compare their efficiency and accuracy in detecting attacks in the automated vehicles domain, as well as compare the performance of algorithms Apriori and FP-growth and Decision tree and Random forest with each other.

The introduction provides a comprehensive background on anomaly detection, emphasising its significance in various domains, including engineering and cybersecurity. It also highlights the relevance of the research in the context of autonomous vehicles, where anomaly detection is crucial for ensuring safety and security. The background is given on the main concepts used in the thesis, and related works demonstrate other similar research in adjacent domains. Methodology is described in detail along with algorithm flow and parameters.

The experimental results demonstrate that machine learning algorithms, particularly Random forest, outperform association rule mining in accurately identifying attacks in the dataset. However, association rule mining has the advantage of not requiring attack data for detection, making it suitable for scenarios where anomalies are not clearly defined. In addition, association rules are capable of detecting attacks in data with a very low presence of attacks.

In conclusion, while machine learning algorithms, specifically Random forest, proved to be effective in predicting specific anomalies like attacks in autonomous vehicle data, association rule mining remains a valuable tool, especially in scenarios where anomalies are ambiguous. The thesis recommends Random forest as the preferred anomaly detection method for this specific dataset, but acknowledges the importance of ARM in certain contexts.

Future research focus could be on further refining anomaly detection methodologies, exploring hybrid approaches, and addressing challenges such as distinguishing between different types of anomalies and optimising detection accuracy, reducing false positives.

KOKKUVÕTE

See lõputöö uurib anomaaliate tuvastamist assotsiatsioonireeglite kaevandamise ja juhendatud masinõppe abil, keskendudes nende rakendamisele autonoomsete sõidukite sensorandmetele rünnakute tuvastamiseks. Eesmärk on võrrelda nende tõhusust ja täpsust rünnakute tuvastamisel automatiseeritud sõidukite valdkonnas, samuti võrrelda Apriori ja FP-growth algoritmide ning otsustuspuu ja juhusliku metsa toimivust omavahel.

Sissejuhatus annab põhjaliku ülevaate anomaaliate tuvastamisest, rõhutades selle olulisust erinevates valdkondades, sealhulgas inseneriteaduses ja küberturvalisuses. Samuti rõhutatakse uurimistöö asjakohasust autonoomsete sõidukite kontekstis, kus anomaaliate tuvastamine on üliloluline ohutuse ja turvalisuse tagamiseks. Tutvustatakse peamisi kontseptsioone, mida lõputöös kasutatakse, ning seotud tööd näitavad teisi sarnaseid uurimusi lähedastel aladel. Metoodika on üksikasjalikult kirjeldatud koos algoritmi voolu ja parameetritega.

Eksperimentaalsed tulemused näitavad, et masinõppe algoritmid, eriti juhuslik mets, ületavad assotsiatsioonireeglite kaevandamise täpsuses rünnakute tuvastamisel andmestikus. Siiski on assotsiatsioonireeglite kaevandamisel eelis, et see ei vaja rünnakute tuvastamiseks andmeid, muutes selle sobivaks olukordades, kus anomaaliad ei ole selgelt määratletud. Lisaks on assotsiatsioonireeglid võimelised tuvastama rünnakuid andmetes, kus rünnakuid esineb väga vähesel määral.

Kokkuvõtteks, kuigi masinõppe algoritmid, eriti juhuslik mets, osutusid efektiivseks konkreetsete anomaaliate, nagu rünnakute, ennustamisel autonoomsete sõidukite andmetes, jääb assotsiatsioonireeglite kaevandamine väärtuslikuks tööriistaks, eriti olukordades, kus anomaaliad on ebamäärased. Lõputöö soovitab juhuslikku metsa eelistatud anomaaliate tuvastamise meetodina selle konkreetse andmestiku jaoks, kuid tunnistab ARM-i olulisust teatud kontekstides.

Tuleviku uurimistöö võiks keskenduda anomaaliate tuvastamise meetodite edasisele täiustamisele, hübriidlähenemiste uurimisele ning väljakutsetele, nagu erinevate anomaaliate eristamine ja tuvastamise täpsuse optimeerimine, vale-positiivsete vähendamine.