

The Open Information Systems Journal

<http://www.bentham.org/open/toisj/>

Editorial Board

S. Ba (University of Connecticut, Storrs, CT, USA)
K. Burke (University of Hawaii, Hilo, HI, USA)
J.M. Carroll (The Pennsylvania State University, University Park, PA, USA)
H.-H. Chen (National Sun Yat-Sen University, Kaohsiung, Taiwan)
Y.-S. Chen (National Taipei University, San Shia, Taiwan)
E.A. Fox (Virginia Tech, Blacksburg, VA, USA)

M. Benaroch (Syracuse University, Syracuse, NY, USA)
Y. Cai (Iowa State University, Ames, IA, USA)
B. Carpentieri (Karl-Franzens University, Graz, Austria)
L. Chen (Osaka Sangyo University, Osaka, Japan)
R.B. Cooper (University of Houston, Texas, USA)
J. Mostafa (Indiana University, Bloomington, IN, USA)

Full list of the board members can be seen at <http://www.bentham.org/open/toisj/EBM.htm>

PAPER TITLE:

“Cloud Data Leakage Prevention mit
Methoden der
automatischen Datenklassifizierung”

By: Anna Hamberger

REFeree'S ASSESSMENT

Please evaluate the enclosed manuscript for each of the following:

(cross as appropriate)

	Excellent	Good	Fair	Poor
Originality:		X		
Technical quality:	X			
Importance in its field:		X		
Style & Overall representation:		X		
Readily Understandable	X			
Strong Focus on Targets	X			
Interesting For a Non-Expert		X		
Adequate Illustrations or Drawings	X			

Overall the paper is rated: (encircle)	10	9	8	7	6	5	4	3	2	1
Potential if revised: (encircle)	10	9	8	7	6	5	4	3	2	1

Where 10 is the highest score

REFeree'S RECOMMENDATIONS

Accept as it is

☐
☒

Accept with minor changes

OTHER SPECIFIC CRITICISMS

Imperfect style

Too long

☐
☐

Accept with revision as attached comments

Reject

References incorrectly presented

Typographical and Grammatical errors

x

REFeree's COMMENTS (continue on another sheet, if necessary)

Inhaltliche Aspekte:

Probleme sind klar beschrieben und die Forschungsfrage ist klar. Relevanz des Themas ist begründet in der Einleitung. Stand der aktuellen Forschung ist erfasst. Eigener Standpunkt ist erkennbar im Ausblick. Neue Perspektive mit automatischer Datenklassifizierung ist vorhanden. Ergebnisse haben Bezug zur Forschungsfrage sind sachlich richtig und nachvollziehbar. Thema wurde gründlich behandelt. Aktuelle Literatur wurde berücksichtigt (automatische Datenklassifizierung). Klassiker wurden berücksichtigt (manuelle Datenklassifizierung). Sekundärquellen wurden vermieden. Internetquellen wurden im Literaturverzeichnis wiedergegeben. Journal wurde eingebunden (Zitat 4). Das Journal ist hochklassig (H-index 129 und Q1) Die Literatur ist auf das Thema fokussiert und die Quellen wurden richtig verstanden.

Methodische Aspekte:

Gliederung ist erkennbar und sinnvoll. Einleitung enthält Übersicht über Arbeit. Einleitung stellt verwendete Methodik dar aber könnte dieses genauer machen.
Ich würde den Ersten Absatz des Ausblicks rausnehmen und mit dem letzten Absatz der Einführung kombinieren. Da dieser genau beschreibt welche Methoden besprochen wurden und es keine „Geheimnisse“ in der Arbeit geben sollte. Es sollte in der Einführung direkt alles aufgelistet werden, was in der Arbeit besprochen wird, um den Leser Zeit zu sparen, ob die Arbeit relevante Informationen für ihn enthält.
Aufbau: Einleitung, Problem, Forschungsmethodik, Lösung, Schlussfolgerung wurde eingehalten.
Kapitel haben sinnvolle Gewichtung. Das Kapitel IV D „Klassifizierung mit künstlicher Intelligenz“ ist zu groß und sollte mehr aufgesplittet werden. Bei IV Methoden Der Automatischen Datenklassifizierung:
Würde ich das Kapitel aufsplitten Punkt C „Klassifizierung mit manueller Definition“ und D „Klassifizierung mit künstlicher Intelligenz“ passen unter den Titel. A „Erkennung von sensiblen Daten“, B „Herausforderungen in der Datenklassifizierung“ und E „Anwendung in der Cloud Security“ sollten unter einem anderen Titel stehen, da sie keine Methoden der Automatischen Datenklassifizierung sind.
Gewählte Forschungsmethodik wurde begründet und korrekt ausgeführt. Behauptungen sind immer belegt. Bezug auf Literatur und Konzept ist gegeben. Es wurde sich mit der Literatur auseinandergesetzt. Vergleichskriterien sind definiert und objektiv. (Genauigkeit von Klassifizierung).
Die Kapitel sind zusammengefasst und verständlich. Es ist erkennbar wann eigener Standpunkt gesetzt wurde und wann zitiert wurde. Die Zentralen Begriffe der Arbeit sind erläutert.

Formale Aspekte:

Kurzfassung ist vorhanden. Abbildungsverzeichnis fehlt, Tabellenverzeichnis fehlt. Literaturverzeichnis vorhanden. Seitenanzahl passt. Seitennummerierung fehlt. Blocksatz passt. Geeignete Grafiken wurden verwendet. Abbildungen sind im Text erläutert. Abbildungen und Tabellen haben eigene Nummer in Bildunterschrift. Abbildungen haben Druckqualität. Schrittgröße in Abbildungen passt. Abbildungen haben Quellenangaben. Absätze sind klar erkennbar. Absätze haben vernünftigen Umfang.
Zitierweise ist einheitlich. Wörtliche Zitate wurden vermieden. Quellenangaben befinden sich im Literaturverzeichnis. Quelle [15] existiert wurde aber nie zitiert. Literaturverzeichnis ist der Reihenfolge nach geordnet. Rechtschreibfehler vorhanden. Typographische Fehler vorhanden. Abkürzungen wurden beim ersten Auftreten ausgeschrieben und die Abkürzung in Klammer gesetzt. Kapitelüberschriften sind kurz und prägnant. Der erste Satz wurde zu Kapitelbeginn formuliert, als gäbe es keine Überschrift. Lesefluss ist gegeben und Formulierungen sind präzise.

Verbesserungsvorschläge: (ich habe den Satz wie er ist immer davor geschrieben damit es leichter in der pdf gefunden werden kann)

Ich würde den Satz umschreiben: Die Funktionen eines DLP-Systems basieren darauf, dass sensible Daten erkannt und in irgendeiner Art und Weise markiert sind => Die Funktionen eines DLP-Systems basieren darauf, dass sensible Daten erkannt und markiert sind.

„Die Klassifizierung funktioniert zum Beispiel nicht, wenn die Daten verschlüsselt oder passwortgeschützt“ sollte „zum Beispiel nicht“ sein.

„Auserdem kann der Ansatz kann bei großen Datenmengen ressourcenintensiv sein“ ein kann zu viel.

„In [32] hat die Einteilung der Daten in zwei Klassen gut funktioniert“ würde ich umschreiben als „In der Methode von Zardari [32] hat die Einteilung der Daten in zwei Klassen gut funktioniert“.

„der mehrere schwache Modelle zu einem starken Modell zusammenzusetzen“ sollte zusammengesetzt sein.

„Kaur, Zandu [36] schlagen für die Klassifizierung von sensiblen Daten eine neue Boosting-Architektur vor.“ Ich würde daraus Kaur und Zandu machen und die Zitierstelle ist hier falsch. Kaur und Zandu sind [32] und nicht [36]. Diese falsche Zitat Nummer ist an späteren Stellen genauso.

Das Zitat [37] für den Wahrscheinlichkeit Bereich stimmt auch nicht und sollte [33] sein. Ich vermute das beim Überarbeiten im Nachhinein die Zitate durcheinander gekommen sind und deshalb die Nummerierung im allgemeinen nicht mehr passt. Dies sollte unbedingt berichtigt werden. Alles über Zitat [43] fehlt im Literaturverzeichnis.

„Das Fingerprinting-Verfahren, das im Kapitel IV-C0b“ sollte wahrscheinlich IV-Cb sein.

„dass vertrauliche Dateien geschreddert, also in Stücke geteilt wird.“ Sollte „in Stücke geteilt werden“ sein.

Confidential comments to the Editor: (if any)

Confidential comments made in strict confidence to the Bentham Science Publishers.

They will not be transmitted directly to the author(s), but may be used to aid revision of the manuscript.

NAME & SIGNATURE OF REFEREE

DATE:

Matthäus Haselmaier

17/12/2023

Field of Expertise of Referee:

Software Engineer

BENTHAM SCIENCE PUBLISHERS