Introduction

Analysing scientific articles is of unique importance in developing critical thinking and deepening the theoretical and practical knowledge acquired during academic training. This analytical process not only allows students to develop a refined critical capacity, but also enables them to identify gaps and research opportunities in the field of decision support systems and Machine Learning, areas that are proving increasingly crucial in solving complex and dynamic problems.

In the context of the subject in question, it is essential that students analyse articles that directly explore the intersection between data mining techniques and emerging cybersecurity challenges. The relevance of this approach lies in the fact that the use of advanced data extraction algorithms and the application of predictive analysis methods can make a significant contribution to the early detection and prevention of cyber threats, providing a robust contribution to the security of information systems.

Emphasis should be placed on critically analysing case studies that illustrate the practical application of these techniques, allowing an in-depth understanding of real contexts where Data Mining approaches have been successfully implemented to mitigate risks and vulnerabilities. In addition, research into new algorithmic approaches constitutes a fertile field for innovation, encouraging students to explore and develop methods that adapt to the rapid transformations of the digital environment.

Finally, the comparison between different threat detection methods, carried out through systematic reviews and comparative analyses, offers a comprehensive perspective that makes it easier to identify the strengths and limitations of each approach. This comparison exercise not only enriches the students' knowledge base, but also prepares them to face real challenges in the implementation and optimisation of decision support systems, where the integration of Machine Learning techniques is proving increasingly indispensable.

In short, the detailed analysis of scientific literature in this context not only fosters solid academic training, but also contributes to the development of essential skills that, in the medium and long term, will have a significant impact on professionals' ability to integrate and apply innovative solutions in the area of cybersecurity.

Recommended Articles Category

Category	Description	Examples of themes
Data Mining Applications in Cybersecurity	Articles demonstrating the practical use of Data Mining techniques (classification, clustering, association rules, etc.) in real cybersecurity problems.	- Use of Random Forest to detect DDoS attacks- Implementation of Neural Networks to detect malware- Comparison of clustering techniques to identify malicious traffic
Intrusion Detection Systems (IDS) and IPS	Studies on the application of Data Mining and Machine Learning algorithms in IDS/IPS, analysing performance and effectiveness.	- Comparison between signature-based and anomaly-based IDS - Performance analysis of IDS with KDD Cup 99 or CICIDS2017
Case Studies with Public Datasets	Empirical analyses using known cybersecurity datasets.	- Evaluation of the CICIDS2017 dataset in attack detection - Comparative study between the KDD Cup 99 and UNSW-NB15 datasets
Use of Deep Learning in Cybersecurity	Articles exploring the application of CNN, LSTM or Neural Networks in malware detection, log analysis and insider threats.	- Convolutional neural networks for malware classification - LSTM for detecting patterns in persistent attacks (APT)
Insider Threats and Behavioural Analysis	Work that uses data mining techniques (such as time series analysis and NLP) to identify suspicious behaviour.	- Analysis of e-mails with NLP for phishing detection- Behavioural analysis of accesses with UEBA (User and Entity Behavior Analytics)

Recommended Sources

- **arXiv:** An open-access repository that gathers preprints of scientific articles in various fields (such as Physics, Maths, Computer Science, among others), allowing early access to innovations in research.
- **B-On:** A platform that aggregates scientific content from Portuguese libraries, offering access to national and international publications, making it easier to search for and consult relevant studies.
- **Google Scholar: An** academic search tool that consolidates the scientific literature available in multiple , allowing you quickly identify studies and check citations and references.
- **Scopus:** a highly credible multidisciplinary database that provides a vast collection of scientific articles and offers tools for bibliometric analyses and the identification of research trends.
- **Web of Science:** An internationally recognised platform for scientific research, providing access high-quality publications and detailed analyses of the impact of studies using specific metrics.
- **IEEE Xplore: A** digital library specialising in engineering, computer science and cybersecurity, which provides articles, conference proceedings and technical standards, and is essential for research into decision support systems and Machine Learning.
- **ACM Digital Library:** Repository of scientific and technical publications in the field of computing and information technology, which contributes to research in Machine Learning, cybersecurity and other related areas.
- **SciELO**: Electronic library focused on the dissemination of scientific journals from Portuguese-speaking and Latin American countries, promoting access to studies in Portuguese and Spanish and contributing to the dissemination of regional knowledge.

Criteria for choosing an article

The choice of a quality scientific article must be based on a set of rigorous criteria that allow for an in-depth understanding of the subject under study, as well as validation of the scientific contribution presented. To this end, it is essential that the article has a detailed methodology, including a detailed description of the datasets used, clarifying their origin, size and relevant characteristics. This transparency not only makes the study replicable, but also allows a critical assessment of the adequacy of the data in relation to the problem being investigated.

In addition, it is imperative that the article clearly describes the algorithms and techniques implemented, justifying their choice and detailing the parameters and configurations used during their application. This methodological explanation helps to highlight the advantages and limitations of the proposed approaches and, where possible, compares them with alternative methods, thus enriching the scientific discussion.

Another key aspect is the presentation of robust experimental results. The article should demonstrate the effectiveness of the methods through quantitative and, if applicable, qualitative data that allows the performance of the algorithms to be verified. The inclusion of comparisons between different approaches, supported by the use of visual elements such as graphs, tables and diagrams, facilitates the interpretation of the results and emphasises the practical implications of the study, promoting in-depth critical analysis.

Finally, the referential framework is an essential criterion. The article should place the research in the context of the state of the art, using relevant and up-to-date references that highlight pioneering work and recent advances in the field. The use of highly credible sources ensures that the study is part of a solid scientific panorama that is coherent with developments in the field of cybersecurity, Data Mining and Machine Learning.

To summarise, the selection of an excellent scientific article involves verifying the existence of a detailed methodology, the presentation of consistent experimental results and the use of relevant references, which together promote the development of critical thinking and contribute to building a robust and up-to-date knowledge base.

Structure for analysing the article

The analysis should be presented on an A4 sheet of paper, printed double-sided, and should include the following essential elements, organised in a clear and structured manner.

The first element to include is the **Identification of the Article**, which should contain the title, the authors, the name of the conference or journal where the study was published and the respective year of publication. This information is essential to situate the article in its academic context and allow it to be easily located and referenced.

Next, an **Article Summary** should be drawn up, consisting of a brief description of the aim of the study and the main contribution made. This summary is intended to provide an overview of the content, allowing the reader to quickly grasp the essence of the research.

The third element is the **Methodology**, which should detail the Data Mining techniques used. It is imperative to describe the datasets used (including their origin, size and characteristics), the algorithms implemented and the evaluation metrics used to assess the performance of the methods applied. This section must be sufficiently detailed to enable the study to be replicated.

Next, the **Results and Discussion** section should be presented, setting out the main conclusions of the study. In this section, it is important to highlight the performance of the models through experimental results and, where possible, to compare different approaches. The use of visual elements, such as graphs and tables, can facilitate the interpretation of the data and strengthen the critical analysis of the results obtained.

Subsequently, the analysis should include the **Limitations** of the study, describing the challenges faced during the implementation of the methods and pointing out potential improvements that could be adopted in future research. This section demonstrates critical rigour and awareness of the restrictions inherent in scientific research.

Finally, the **Conclusion** should summarise a critical opinion on the relevance of the study and its practical application, summarising the contributions and suggesting possible lines of future research. In addition, the analysis should include a **References** section, listing all the sources cited, ensuring the transparency and credibility of the work presented.


ESCOLA NAVAL
Infantes bien faire

Article evaluation

The assessment of the article, within the framework of this assignment, is responsible for checking not only the quality of the critical and analytical content, but also the student's ability to effectively communicate the results of their analysis. Thus, the assessment will consist of two fundamental components: the written analysis of the article and its oral presentation, both of which play a crucial role in the overall assessment of the student's performance.

1. Article Analysis

The written component should reflect an in-depth treatment of the selected article, demonstrating critical thinking and a robust analytical capacity. The following aspects will be considered in the assessment:

- **Clarity and Objectivity**: The text must be well-structured, cohesive and written clearly and precisely, without ambiguity or redundancy.
- **Depth of Analysis**: The ability to critically examine the premises, methodology, results and conclusions of the article, demonstrating a mastery of the subject.
- **Argumentative richness**: Students should present well-founded arguments, based on academic references and relevant examples.
- **Methodological rigour**: The discussion should demonstrate a solid understanding of the article's methodological approach, highlighting any limitations or potential improvements.

2. Oral presentation in class

In addition to the written analysis, the oral presentation is an essential element of the assessment, lasting a maximum of **20 minutes**. This presentation should be structured to ensure maximum clarity and impact in the transmission of information. The assessment will focus on the following aspects:

2.1 Structure and Clarity of the Presentation

- The presentation should have a **clear beginning**, where the article is contextualised and the objectives of the analysis are defined.

- The **development** should emphasise the key points of the article, ensuring an accessible and well-founded explanation.

2.2 Quality of Analysis and Interpretation

- The ability to identify and explain the most relevant aspects of the article.
- The depth of the methodological analysis, highlighting any limitations or implications of the results.

2.3 Communication and Oral Expression

- **Clarity and objectivity** in the exposition of ideas.
- **Appropriate use of technical and scientific language**, without excessive unnecessary jargon.
- **Posture and diction**, ensuring that communication is fluid and engaging.
- **Visual support** (slides or other materials) that is well organised and complements the presentation.

3. Evaluation Criteria

The final assessment will be weighted on the basis of the following criteria:

Criteria	Subcriteria	Weight (%)
Written analysis	Clarity, depth and structure of the text	30%
Oral Presentation	Structure and clarity of the presentation	20%
	Quality of analysis and interpretation	20%
	Communication and oral expression	15%
Interaction and Debate	Answering questions and encouraging participation	15%

Examples of Recommended Articles

1. **Piskozub, A.** (2024). *Data mining for threat detection in Active Directory*. ResearchGate. https://www.researchgate.net/publication/388886846
2. **Kumar, P., Kushwaha, C., Sethi, D., Ghosh, D., G Gupta, P.** (2023). Investigating the performance of multivariate LSTM models to predict the occurrence of Distributed Denial of Service (DDoS) attack. *PloS ONE*, 18(7). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0313930
3. **D'Antonio, S., G Uccello, F.** (2022). Artificial intelligence applications in healthcare security. In *Advances in Cybersecurity Technologies* (pp. 120-134). Springer. https://link.springer.com/chapter/10.1007/978-3-031-70775-9_9
4. **Guru, A., Gopal, A. V., G Bandarupalli, S. S. B.** (2023). Uncovering threats: Data mining techniques for cybersecurity. *Risk Assessment and Management Journal*, 15(2), 34-45. https://ramd.reapress.com/journal/article/view/52
5. **Wu, E. H. K., G Lin, Y. D.** (2021). TRACE: Relationship analysis and causal factor extraction in cyber threat intelligence reports. *IEEE Transactions on Dependable and Secure Computing*, 18(9), 874-889. https://ieeexplore.ieee.org/abstract/document/10851819
6. **Velasquez, J. D., Pant, M., Pan, J. S., G Snasel, V**. (2022). On the fuzzy entropy and the rankability of data. *SSRN Electronic Journal.* https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5134638
7. **Song, D.** (2023). Penetration testing automation with inverse soft-Q learning: An imitation learning method. *DiVA* Portal. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1935147
8. **Zhao, X., Leng, X., Wang, L., G Liu, Y.** (2023). Efficient anomaly detection in tabular cybersecurity data using large language models. *Scientific Reports*, 13(4), 567-578. https://www.nature.com/articles/s41598-025-88050-z
9. **Papoutsis, A., G Panagiotou, P.** (2020). AI-based holistic framework for cyber threat intelligence management. *IEEE Transactions on Information Forensics and Security*, 15(7), 1024-1039. https://ieeexplore.ieee.org/abstract/document/10851288

10. Bollmann, C. A., Tummala, M., C McEachen, J. C. (2021). Resilient real-time network anomaly detection using novel non-parametric statistical tests. *Computers & Security, 102, 10214C. doi:https://doi.org/10.101C/j.cose.2020.10214*

11. Gibert, D., Mateu, C., Planes, J., C Marques-Silva, J. (2021). Auditing static machine learning anti-Malware tools against metamorphic attacks. *Computers & Security, 102, 10215S. doi:https://doi.org/10.101C/j.cose.2020.10215S*

12. Krumay, B., Bernroider, E. W. N., Walser, R. (2018). *Evaluation of Cybersecurity Management Controls and Metrics of Critical Infrastructures: A Literature Review Considering the NIST Cybersecurity Framework, Cham.*

13. Lin, W.-C., Ke, S.-W., C Tsai, C.-F. (2015). CANN: An intrusion detection system based on combining cluster centres and nearest neighbors. *Knowledge-Based Systems, 78, 13-21. doi:https://doi.org/10.101C/j.knosys.2015.01.00S*

14. Mitchell, R., C Chen, I.-R. (2014). A survey of intrusion detection techniques for cyber-physical systems. *ACM Comput. Surv., 4C(4), Article 55. doi:10.1145/254204S*

15. Casas, P., Mazel, J., C Owezarski, P. (2012). Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge. *Computer Communications, 35(7), 772-783. doi:https://doi.org/10.101C/j.comcom.2012.01.01C*

16. García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., C Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers C Securit

17. Data Mining for Cyber Security, V.Chandois *et al.*, in Data Warehousing and Data Mining Techniques for Computer Security, Springer, 2006.

18. Data mining methods for anomaly detection KDD-2005 workshop report, Margineantu *et al.,* ACM SIGKDD Explorations Newsletter, Volume 7 Issue 2, December 2005.

19. On the efficacy of data mining for security applications, Ted E. Senator, ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics -CSI-KDD '09, 2009.

20. Metrics for mitigating cybersecurity threats to networks, IEEE Internet Computing, 14, 1, Jan-Feb 2010.
21. A Combined Fusion and Data Mining Framework for the Detection of Botnets, Kiayias *et al.*, Conference For Homeland Security, 2009. CATCH '09. Cybersecurity Applications C Technology, March 2009
22. A study of Spam Detection Algorithms on Social Media Networks, Jacob Soman Saini, International Conference on Computational Intelligence, Cyber Security, and Computational Models, Coimbatore, India, December 2013.
23. Comparative Study of Two- and Multi-Class-Classification-Based Detection of Malicious Executables Using Soft Computing Techniques on Exhaustive Feature Set. Shina Sheen, R. Karthik and R. Anitha; International Conference on Computational Intelligence, Cyber Security, and Computational Models, Coimbatore, India, December 2013
24. Botnets: A Study and Analysis, G. Kirubavathi and R. Anitha, International Conference on Computational Intelligence, Cyber Security, and Computational Models, Coimbatore, India, December 2013
25. The VoIP intrusion detection through a LVQ-based neural network, Zheng Lu ; Taoxin Peng, International Conference for Internet Technology and Secured Transactions, 2009. ICITST 2009.
26. Detection of applications within encrypted tunnels using packet size distributions, Mujtaba,G.,Parish, D.J., International Conference for Internet Technology and Secured Transactions, 2009. ICITST 2009.
27. Email classification: Solution with back propagation technique, Ayodele et al. International Conference for Internet Technology and Secured Transactions, 2009. ICITST 2009.
28. Malware detection using statistical analysis of byte-level file content, Tabish et al., CSI-KDD '09 Proceedings of the ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics, 2009