

BUCHAREST UNIVERSITY OF ECONOMIC STUDIES
Council for Doctoral University Studies
Doctoral School of Economic Informatics

**ADVANCED DATA ANALYSIS AND VISUALIZATION
TECHNIQUES FOR E-LEARNING**

Marian STAN

Doctoral supervisor: Prof. Ph.D. ION SMEUREANU

Bucharest, 2025

TABLE OF CONTENTS

Chapter I - Introduction.....	1
1.1 Importance of the research topic.....	1
1.2 Context and relevance of e-learning.....	2
1.3 The necessity of advanced data analysis and visualization techniques.....	3
1.4 Thesis objectives.....	4
1.5 Main contributions.....	5
1.6 Ontological, epistemological and axiological foundations of the thesis.....	6
1.7 Structure of the thesis.....	8
Chapter II - Current state of knowledge.....	9
2.1 Bibliographic research methodology.....	9
2.2 Results of the bibliometric analysis.....	12
2.3 Co-occurrence analysis of keywords.....	14
2.4 Thematic approach of the literature review.....	18
2.5 Conclusions and implications for research.....	21
Chapter III - Methodological dimensions.....	23
3.1 Research objectives and research questions.....	23
3.2 Research design.....	24
3.3 Documentary analysis.....	25
3.4 Multiple case study.....	25
3.5 Literature selection.....	26
3.6 Datasets and quantitative analysis techniques.....	26
3.7 Development of the application prototype (Python and Streamlit).....	28
3.8 Methodological limitations.....	30
Chapter IV - Data preprocessing and cleaning in e-learning.....	31
4.1 Data sources in e-learning.....	31
4.2 Data collection methods.....	32
4.3 Data preprocessing and cleaning.....	33
Chapter V – Analytical maturity, organizational model and case studies in e-learning.....	38
5.1 The Davenport-Harris analytical maturity model.....	38

5.2 Organizational structure and data governance.....	41
5.3 Key Performance Indicators (KPIs) as a measure of maturity.....	43
5.4 Case studies.....	44
5.5 Synthesis of main findings.....	58
5.6 Contribution to thesis objectives.....	59
Chapter VI – Advanced data analysis techniques for e-learning.....	61
6.1 Introduction.....	61
6.2 Predictive analysis and classification in e-learning.....	62
6.3 Unsupervised learning techniques.....	75
6.4 Interaction and sequential data analysis.....	81
6.5 Text analysis techniques and Generative AI.....	91
Chapter VII – Advanced data visualization techniques for e-learning.....	97
7.1 Introduction to data visualization for e-learning.....	97
7.2 Libraries and tools for data visualization in Python.....	97
7.3 Data visualization for dropout analysis.....	99
7.4 Visualization of study material difficulty level.....	100
7.5 Visualization of clusters in e-learning.....	104
7.6 Visualization of atypical behaviors in e-learning.....	106
7.7 Visualization of social networks in e-learning.....	108
7.8 Visualization of sequential data mining, learning paths and clickstream sequences.....	110
7.9 Visualization of sentiment analysis.....	116
7.10 Dashboards for data visualization.....	121
Chapter VIII – Data analysis and visualization platform for e-learning.....	125
8.1 Motivation for choosing the general architecture and programming language.....	125
8.2 Application functionalities by modules.....	125
8.3 Prototype implementation.....	129
8.4 Prototype functionalities.....	130
8.5 Automated narrative with OpenAI API.....	135

8.6 Application evaluation.....	135
8.7 Technical and functional limitations of the application.....	136
8.8 Future development directions.....	136
8.9 Application contributions to thesis objectives.....	137
Chapter IX - Contributions and future research directions.....	138
9.1 Modular architecture for educational data analysis.....	138
9.2 Customized dashboards for different educational roles.....	138
9.3 Integration of GenAI for explanatory automated narratives.....	139
9.4 Validation of contributions and results.....	140
9.5 Limitations of the current research.....	140
9.6 Future research directions.....	141
9.7 Conclusions.....	141
Bibliography.....	143
Appendix A – Application source code.....	155
Appendix B – List of abbreviations used in the thesis.....	196
Appendix C – List of figures and tables.....	200
Appendix D – Description of datasets used.....	203
Appendix E – List of Python libraries and their versions.....	207
Appendix F – List of published works.....	211

Keywords: data analysis, data visualization, e-learning.

SUMMARY

The doctoral thesis addresses a highly relevant topic in the context of the digitalization of education, exploring the application of advanced data analysis and visualization techniques in the field of e-learning. The study aims to contribute to the optimization of online educational processes by developing innovative solutions for the analysis of educational data.

The main objectives include examining the current state of knowledge in the field, exploring the capabilities of e-learning platforms through the Davenport-Harris analytical maturity

model, proposing an organizational model to promote advanced data analysis, and developing a prototype application using Python and Streamlit.

The research methodology is mixed, combining documentary analysis, multiple case studies, and the development of an application prototype. The study investigates techniques such as dropout prediction, clustering analysis, anomaly detection, social network analysis, sequential data mining, and sentiment analysis, applying these methods to public datasets (Open University Learning Analytics Dataset or Gartner Peer Insights).

The main contributions consist of proposing an e-learning-specific organizational model oriented toward advanced analytics, developing a conceptual framework for performance measurement through KPIs, applying innovative methods of analysis and visualization exemplified through Python code, and creating a prototype application that integrates functionalities for analysis, visualization, and generative artificial intelligence (GenAI) for the automatic interpretation of results.

The structure of the thesis comprises nine chapters: introduction, state of the art, methodology, data preprocessing, analytical maturity, and case studies (Coursera, Udemy, LinkedIn Learning, The Open University), analysis techniques, visualization techniques, platform development, and contributions with future directions.

The results highlight the potential of advanced data analysis and visualization techniques to transform the understanding and optimization of learning processes in digital environments. The developed prototype provides functionalities for dropout prediction, clustering analysis, outlier detection, interactive dashboards, and automated narrative generation with the OpenAI API, marking a transition toward prescriptive analytics in e-learning.

The thesis opens new research directions and provides a solid foundation for developing more efficient and personalized e-learning solutions, making a significant contribution to understanding the application of data analytics in online education.