

Victoria L. Prince

PhD | Sr. Statistician / Data Scientist.

Research Computing Services, Harvard Business School, Boston, MA 02163, USA

L_tory@hotmail.com

PRESENTATION ON BIG DATA

Abstract This presentation explores the diverse techniques and tools integral to big data analysis, a field pivotal in transforming vast datasets into meaningful insights. In the age of information, where data grows exponentially, efficient analysis methods are paramount. This work delves into various analytical techniques, from machine learning to predictive analytics, and examines key tools like Hadoop and Apache Spark. The objective is to provide a comprehensive understanding of current methodologies and tools, their applications, and the challenges they face, paving the way for future advancements in big data analysis.

Ключові слова: big data analysis; big data analysis techniques.

1. INTRODUCTION

Big data, characterized by its volume, velocity, and variety, presents a unique set of challenges and opportunities. This presentation introduces the concept of big data and underscores its growing relevance across industries. It aims to dissect the complexities of big data analysis, highlighting the need for robust techniques and tools to handle this data deluge. The objectives include a comparative analysis of various big data analytical methods, an

exploration of the tools employed in this field, and an understanding of the challenges and future directions of big data analytics.

2. THE EVOLUTION OF BIG DATA ANALYSIS

Big data analysis has evolved significantly from traditional data processing methods. Early data analysis was limited by computational power and storage capabilities, focusing mainly on structured data. The advent of the internet and digital technologies catalyzed an exponential increase in data volume, variety, and velocity, leading to the emergence of big data. Key technological advancements such as distributed computing and advances in artificial intelligence have been instrumental in this evolution. The development of platforms like Hadoop and the adoption of machine learning algorithms have transformed data analysis, enabling the processing and analysis of large, unstructured datasets in real-time.

3. BIG DATA ANALYSIS TECHNIQUES

Big data analysis encompasses a range of techniques. Machine learning, a subset of AI, empowers systems to learn and improve from experience, making it ideal for pattern recognition in large datasets. Data mining involves extracting valuable information from large sets of data, revealing patterns and relationships. Predictive analytics uses statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. Each technique offers unique advantages: machine learning excels in handling unstructured data, data mining is effective for discovering hidden patterns, and predictive analytics is essential for forecasting.

4. TOOLS FOR BIG DATA ANALYSIS

Various tools have emerged to facilitate big data analysis. Hadoop, an open-source framework, allows for distributed storage and processing of large data sets across computer clusters. Apache Spark, another open-source distributed computing system, provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. NoSQL databases, like MongoDB, cater to the storage needs of big data by providing high performance and scalability. These tools differ in processing capabilities, speed, and complexity; Hadoop is renowned for its high storage capacity and fault tolerance, Spark for its speed in data processing, and NoSQL databases for their flexibility in handling unstructured data.

5. CASE STUDIES

In healthcare, big data analysis has enabled predictive modeling for patient outcomes, improving treatment plans. In finance, it's used for risk analysis and algorithmic trading, leading to more informed investment decisions. Retail companies leverage big data for personalized marketing and inventory management. These cases illustrate the transformative impact of big data analysis across sectors, driving efficiency, innovation, and strategic decision-making.

6. CHALLENGES AND FUTURE DIRECTIONS

Despite its potential, big data analysis faces significant challenges. Data privacy and security are paramount concerns, especially with the increasing sensitivity of information. Ensuring data quality and integrating diverse data sources also pose significant hurdles. Future directions may include advancements in AI and machine learning algorithms, greater focus on data privacy, and the development of more sophisticated tools for real-time analysis.

7. CONCLUSION

This presentation underscores the importance of big data analysis in the modern data-driven world. By examining various techniques and tools, it highlights the ongoing evolution

in this field. The continuous development of these methodologies is crucial for turning vast, complex datasets into actionable insights.

MINISTRY OF EDUCATION
AND SCIENCE OF UKRAINE

NATIONAL UNIVERSITY
OF LIFE AND ENVIRONMENTAL
SCIENCES OF UKRAINE

FACULTY OF INFORMATION
TECHNOLOGY

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ

PROCEEDINGS

XI International scientific
conference

**GLOBAL AND
REGIONAL PROBLEMS OF
INFORMATIZATION IN
SOCIETY AND
NATURE USING
'2023**

15-16 November 2023

Kyiv, NULES of Ukraine

Kyiv 2023

МАТЕРІАЛИ

XI Міжнародної науково-практичної
конференції

**ГЛОБАЛЬНІ ТА
РЕГІОНАЛЬНІ ПРОБЛЕМИ
ІНФОРМАТИЗАЦІЇ В
СУСПІЛЬСТВІ І
ПРИРОДОКОРИСТУВАННІ
'2023**

15-16 листопада 2023 року

Київ, НУБіП України

Київ 2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

МАТЕРІАЛИ

XI Міжнародної науково-практичної конференції

ГЛОБАЛЬНІ ТА РЕГІОНАЛЬНІ ПРОБЛЕМИ ІНФОРМАТИЗАЦІЇ В СУСПІЛЬСТВІ І ПРИРОДОКОРИСТУВАННІ '2023

15-16 листопада 2023 року

Київ, НУБіП України

Київ 2023

УДК 004

Рекомендовано до друку вченою радою факультету інформаційних технологій Національного університету біоресурсів і природокористування України (протокол № 4 від 20.11.2023)

Укладач: к.е.н., доцент Харченко В.В.

Збірник матеріалів XI Міжнародної науково-практичної конференції "Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні '2023", 15-16 листопада 2023 року, НУБіП України, К. НУБіП України, 2023. 117 с.

Відповідальність за зміст публікацій несуть автори.

© Національний університет біоресурсів
і природокористування України, 2023