International Journal of Education, Business and Economics Research (IJEBER)

ISSN: 2583-3006

Vol. 4, Issue.3, May-June 2024, pp 238-256

To cite this article: Sabiha Jabeen Anwar, Hajira Banu, Ruhiya Sarwar Nazneen, Dr. Tariq Zubair and Dr. Mohammed Maqsood Ali (2024). Data Science For The Next Millennium: A Ubiquitous Review. International Journal of Education, Business and Economics Research (IJEBER) 4 (3): 238-256

DATA SCIENCE FOR THE NEXT MILLENNIUM: A UBIQUITOUS REVIEW

Sabiha Jabeen Anwar[1], Hajira Banu[2], Ruhiya Sarwar Nazneen[3], Dr. Tariq Zubair[4] and Dr. Mohammed Maqsood Ali[5]

[123]Lecturers, Department of Computer Sciences,
Applied College, Jazan University, KSA
[45]Assistant Professor, Department of Business Administration,
Applied College, Jazan University, KSA

https://doi.org/10.59822/IJEBER.2024.4317

ABSTRACT

Simple algorithms and analytical tools are not capable of processing the complex data into meaning insights. Data science receives widespread attention in all the aspects of business operations and strategies and moves into the next millennium. Organizations use it to predict, classify, detect, communicate, and analyze complex data for business planning and decision makings. This paper provides a comprehensive review of data science, advanced tools, and its applications in the aspects of business sectors. Articles previously published for periods of 6 years (2018-2023) from various scientific databases (EBSCO, Web of Science, Scopus, Emerald, ProQuest, Wiley, Taylor and Francis) are selected for the study. Results of this review revealed that data science is the intersection of various fields such as computer sciences, business, operation research, engineering, machine learning and automation. The main purpose of data science is to solve problem (s) and develop insights from the extracted information and knowledge of data. Majority of research studies conducted in the context of education and healthcare across 20 countries. This review energizes to all types of business practitioners and academicians to advance data science for decision makings and future research directions.

KEYWORDS:- Data science, Machine Learning, Artificial Intelligence, Data Mining.

1. INTRODUCTION

Data is ubiquitous. The term ‘data’ refers to computer information that is either transmitted or stored in numerous forms (text, numbers, images, videos, spreadsheets, social media, and social

networks) for the purpose of analysis to achieve organizational goals. Previously, organizations prepare data from different sources (text file, images, multimedia etc) in small size by using simple algorithms or analytical tools to predict, classify, communicate, and analyze complex data for business planning and decision makings. In recent times, data science has received enormous attention (Dongwoo Chae, 2020) and emerged as a new discipline (Lin Wang, 2018) that increases an ability to collect and process large amounts of data to extract business insights. Organizations use the notion of data science to process complex raw data into meaningful insights by using advanced algorithms or analytical tools, for example machine leaning, Artificial Intelligence, Neural networks, Internet of Things, data mining, big data and so on, in order to achieve their major goals. Thus, data science gathers, organizes and maintains data knowledge that can be used for experiments and prediction (SanketMantri, 2018), forecasting, market differentiations, and customer or user experiences. In addition, data science not only produces a holistic view but also provide solutions for complex phenomena by using various theories and methods.

The notion of data science has been utilizing in all aspects of businesses (healthcare, educations, manufacturing, banking, insurance, tourism, communication, transportation and so on) and achieving tremendous growth. In addition, data science is high demand and transforming businesses to make critical decision(s) is thus a focus of major research. The evidence from the reviewed articles suggests lack of concise framework for understanding the concepts and its application in businesses. Therefore, the main purposes of this ubiquitous review is to asses and synthesizes the data science concepts, methods and its applications in all aspects of businesses and provide future research directions to businesses practitioners and academicians to accelerate the discipline of data science. In addition, this review also energizes researchers by a need for a comprehensive understanding of the concepts of data science and its characteristics, uses, and tools in all aspects of business operations globally.

The rest of the paper comprises of five sections. Section 2 outlines the literature of previously published studies. Research method has discussed in the section 3. Section 4 focuses on the findings of the study and finally section 5 highlights the discussions and conclusions.

2. BACKGROUND (see SLR data science pdf)

Today, data has been utilizing in a wide variety of businesses or industries. In businesses, data for example inventory, sales, purchases, customer satisfaction, competitors, customer care and support, are acquired and stored in clouds or computing clusters. The stored data in the clouds or computers are investigated, organized and maintained for meaningful interpretations is called data science. The notion of data science is a relatively young discipline and a fast-paced filed and change is inevitable**.** Numerous studies have attempted to define data science in different ways**.** Data science is a set of fundamental principles **(Provost and Fawcett, 2013),** a part of business intelligence **(Larson and Chang, 2016)** aimed at addressing big data problems **(Song and Zhu, 2016).** Moreover, data science is the systematic study of the organization, properties, and analysis of data **(Dhar, 2013),** an interdisciplinary involving statistics, data mining, machine learning and data analytics **(Georgeetal, 2016),** is a combination of statistics, computer sciences and information design**(Shum et al., 2013)** including not only statistics but also computer science (hardware and software engineering**(Diggle, 2015).** In addition, data science is a combination of techniques, knowledge and skills applied to data to identify hidden knowledge that can be used to guide teams in making important decisions **(Parks, 2017).** Furthermore, it process of multi-stage knowledge discovery in which useful knowledge is extracted from a raw, often impure collection in a specific context **(Turkay et al., 2018).**

Data Science, On the other hand, is a cyclical process of capturing the business needs and acquiring the relevant data, storage, security, privacy, preparation, pre-processing, analytics; generating and communicating insights; and finally, actuating these actionable insights **(Jägare 2019).**

Organizations store complex raw data in structured format to extract valuable knowledge to improve their day to-day activities. Therefore, every aspects of life have changed due to complex raw data.

Some of the benefits of data science in businesses are: gaining customer insights, increasing security, predicting future market trends, measuring customer satisfactions, forecasting budget and sales, handling operations, and so on. This review studied the previously published research articles conducted by the various research scholars in the field of data sciences across the globe to understand the concepts, methods and application of data science. Furthermore, determine the trends and study area of research across the globe**.**

An attempt has been made in the past to conduct literature review systematically, realistically and critically to identify various components of data sciences from the different point of view. This ubiquitous review will be the first study to identify the components of data science from the business point of view.

3. RESEARCH METHODS (see SLR data science pdf)

To achieve the objectives of this review, first, a method which consist of research questions, relevant studies, select studies, chart the data, summarize and present findings (Arksey& O'Malley, 2005) are adopted. Second, researchers identified literature from multiple databases according to data science (EBSCO, Taylor and Francis, Wiley, Sage, Emerald and Google Scholars) in all aspects of business globally (see table 1). The search criterion was based on "data science" refined through an iterative process in all aspects of business across the globe for a period of ten years (2012-2022). Then, a total of 105 published articles in English language were included from the above databases. The selected published journal articles were thoroughly screened to attend the purpose of review by excluding if there was no explicit reference to data science.

Table 1: Iterative Review process

S.No	Database Sources	Number of articles
1	EBSCO	15
2	Taylor and Francis	20
3	Wiley	15
4	Sage	12
5	Emerald	8
6	Google Scholar	5
	Total	**105**

4. FINDINGS

This study reviewed the previous research work conducted by the most eminent research scholars and practitioners towards the data sciences. Table 1 reveals the summary of previous research studies during the year 2018 to 2023.

Table 1: Summary of Previous studies

Author	Years	Objectives	Findings	Contexts	Country	Tools	Type of study
Kirby McMaster et al.	2018	To compare and examine the word rates in Data Analytics and Data Science documents to determine which concepts are most often use.	One difference is that the words problem and solution had Top 25 word rates for Data Science, but not for Data Analytics. Data Analytics is more focused on exploratory concerns (searching for patterns in data) while Data Science retains more of the classical inferential activities (sample data to draw conclusions about populations).	Education	USA	ANOVA, multiple regression, and cross tabulation	Empirical
Lin Wang	2018	To examine how data science relates to information sciences in schools of library and information science	Data science and information science are twin disciplines by nature. The mission, task and nature of data science are consistent with those of information science.	Education	China	---	Conceptual
Pedro Galeano and Daniel Peña	2019	To analyze how Big Data world is producing in statistical data analysis.	Statistics (1) collect the data, by sample surveys or designing experiments; (2) describe the data, by plots and summary statistics, and selecting a possible model or a set of models; (3) estimate the model parameters, by maximum likelihood or Bayesian estimation, and making validation of the model or model selection; and (4) interpret the result	General	Spain	Bayesian Theory Machine Learning	Conceptual
Adam Loy et al.	2019	To develop, implement, and evaluate a series of tutorials and case studies that highlight fundamental tools of data science (visualization, data manipulation , and database usage) in statistics curricula.	Fundamental tools of data science into the statistics curriculum includes nine tutorials : (1) tidy data (2) data wrangling, (3) merging data, (4) data visualization, (5) working with shape files, (6) working with dates and times, (7) working with text data , (8) data scraping, and (9) classification and regression trees.	Education	USA	R and R Markdown. Stat 101 Toolkit	Review
Robert Lucero et al.	2019	To enhance the safety of hospitalized older adults by reducing ICs through an effective learning health system.	Applying a state-of-the-art practice-based data science approach to identify risk factors of ICs (e.g., falls) from structured (i.e., nursing, clinical, administrative) and unstructured or text (i.e.,	EPIDEMIOLOGY	USA	Machine Learning and Text-Mining	Review

			registered nurses progress notes) data.				
Sergey V. Novikov	2020	To explore the role of Data Science and Big Data technology in the modern digital economy.	The role of Big Data technology is to be a liquid product and a necessary condition to increase the profitability of enterprises through personalized customer service and predictive analytic	General	Russia	CRISP-DM data cycle	Empirical
Mujthaba et.al	2020	To study data science techniques, tools and predictions	*Data science tools: Excel, R Programming Language, Tableau Public, Python, SAS, Apache Spark. *Data processing tools: Hadoop, Cassandra, Cloudera, Flink, Qubole, Statwing, Storm and couchDB. *Data Analysis tools: Rapidminer, Qlikview, Excel, SAS, Python, Tableau public R and Splunk	General	India	Python, SAS, R, Excel	Conceptual
Martin G. Tol sgaar et al.	2020	To explore applications of data science and ML in HPE literature	Integration of data science in the field of HPE is in danger of becoming technically driven and narrowly focused in its approach to teaching, learning and assessment.	Health Professions Education	Switzerland	Computer-based algorithms , Artificial Learning and Machine Learning (ML)	Review
VerónicaCuellо et al. Need to check	2020	To introduce a comprehensive end-to-end solution aimed at enabling the application of state-of-the-art Data Science and Analytic methodologies to a food science related problem	The problem refers to the automation of load, homogenization, complex processing and real-time accessibility to low molecular-weight relators (LMWGs) data to gain insights into their assembly behavior. The complexity of those databases, and the errors caused by manual data entry, can interfere with the analysis and visualization of relations and patterns, limiting the utility of the experimental work.	Health	Argentina	Big Data, SQL	Empirical
Eva N. Hamulyák et al.	2020	To determine how the representation of women's health has changed in clinical studies over the course of 70 years.	Study observed that a word in a BMJ research article was 'woman' or 'women' increased by an annual factor of 1.023, but this rate of increase varied by clinical specialty with some showing little or no change	Health	UK	Text mining tools. General linear, additive and segmented	Empirical

						regressio n models	
Erickson, Tim and Chen, Ernest	2020	To describes a short module for introducing data science to senior school students or other data-science beginners	Students use CODAP to do their work	Education	USA	CODAP, a free, web-based data analysis platform	Conceptual
Huadonget.al.	2020	to study the complex interaction between physical and social phenomena using data science	An understanding of the content of the datasets collected and pre-processed by the Big Earth data science platform: artificial intelligence, Internet of Things, and Digital Twins and Metaphors			Data Mining, Machine Learning, Collectiv e Intelligen ce and Data Visualiza tion	Review
Maria Jose´ Sousa et al.	2021	To compare the strategies of companies with data science practices and methodologies and the data specificities/varia bles that can influence the definition of a data science strategy in pharma companies	This study identified the following data variables that can influence the strategies of pharma companies. overwhelming volume, managing unstructured data, data quality, availability of data, access rights to data, data ownership issues, cost of data, lack of pre-processing facilities, lack of technology, shortage of talent/skills, privacy concerns and regulatory risks, security, and difficulties of data portability regarding companies with a data science strategy and companies without a data science strategy	Healthcare	UK	a covarian ce analysis and Shapiro–Wilk test	Empirical
João Victor da Silva Guerra et al.	2021	To develop a Python (pyKVFinder) package to detect and characterize cavities in bio molecular structures for data science and automated pipelines.	pyKVFinderefciently detects cavities in bio molecular structures and computes their volume, area, depth and hydropathy, storing these cavity properties in NumPy arrays. pyKVFinder can be integrated with data science packages such as matplotlib, NGL Viewer, SciPy and Jupyter notebook and it can also be integrated with machine learning and 3D visualization in automated pipelines.	Healthcare		pyKVFin der, SciPy and Jupyter notebook	Empirical
VarlamKutate	2021	To study group	Pooled testing has been	Healthcare	USA	Pooled	Empirical

ladze and Ekaterina Seregina		testing from a data science perspective for extensive empirical comparison of group testing techniques based on simulated data	successfully employed against a number of diseases and found effective against SARS-CoV-2 for more than 70 years. Pooling can be a viable method, and compatible with testing kits such as RT-qPCR. Group testing decline infection rates, claimed by FDA (Food and Drug Administration, US).			testing RT-qPCR Test Non-Adaptive Group Testing	
Thomas Donoghue et al.	2021	To develop new courses and programs to meet the growing demand for data science education	In teaching data science as a course for (1) Conceptualizing data science as creative problem solving, with a focus on project-based learning, (2) prioritizing practical application, teaching and using standardized tools and best practices, and (3) scaling education through coursework that enables hands-on and classroom learning in a large-enrollment course.	Education	USA	JSON, CSV, and XML, etc.	Review
Thomas Donoghue et al.	2021	To study new courses and programs to meet the growing demand for data science education.	In teaching data science as a course for (1) conceptualizing data science as creative problem solving, with a focus on project-based learning, (20 prioritizing practical application, teaching and using standardized tools and best practices, and (3) scaling education through coursework that enables hands-on and classroom learning in a large-enrollment course.	Education	USA	JSON, CSV, XML,	Review
Aimee Schwab-McCoy et al.	2021	To investigate the challenges that are facing by faculty members towards data science course teaching in the classrooms.	Study revealed that one of the biggest challenge is teaching computing skills to a diverse audience with varying preparation in introductory data science (albeit fewer computing topics than statistics topics).	Education	USA	Java, Julia, Python, R, SQL	Empirical
Brian Kim and GrahaHenke	2021	To discuss the advantages of using cloud computing to solve the coding issues in different languages.	Using user-friendly Jupyter notebooks along with the interactive capabilities possible through Binder, we provide introductory Python and SQL material that students can access without downloading anything. This lets students to get started with coding right away	Education	USA	Jupyter notebooks with Binder, Python and SQLite	Conceptual

			without getting frustrated figuring out what to install.				
Nathan C Emery et al.	2021	To investigate how data science is taught by biological and environmental science instructors in higher education	Instructors use, teach, and view data management, analysis, and visualization as important data science skills. Coding, modeling, and reproducibility were less valued by the instructors, although this differed according to institution type and career stage.	Education	UK	R (R Core Team for statistical analysis).	Empirical
Kumar Das et al.	2021	To gain insights into SARS-CoV-2 and the outbreak of COVID-19 in order to forecast, diagnose and come up with a drug to tackle the virus	The majority of the COVID-19 related research relies on genomics and proteomics data of SARS-CoV-2 and other coronaviruses more than transcriptomic and metabolomics data	Healthcare	Italy	Phylogenetic analysis and clinical imaging	Empirical
Rasha M. Abd El-Aziz et al.	2021	To introduce an effective data science technique for IoT supported healthcare monitoring system with the rapid adoption of cloud computing.	The sensitivity of BS-DNN is high compared with those of the other classifiers. The specificity of BS-DNN is also high when compared with those of the other classifiers. Similarly, there call and scale of BS-DNN classifier are also high when compared with those of the other classifiers.	Healthcare	Saudi Arabia	Improved Pigeon Optimization (IPO) algorithm and Deep Neural Network (BS-DNN)	Empirical
Ward et al.	2021	To review major concepts in SDS (surgical data science) and AI as applied to surgical education and surgical oncology.	AI in surgery. AI for surgical performance augmentation and education. Automated phase and instrument recognition.	Education	USA	Artificial intelligence (AI)	Review
Leonardo Carvalho et al.	2021	To examines the data analytics applications used for investigating flight delays	Classification Cluster Analysis Machine Learning Network Analysis Pattern Mining Regression Statistical Analysis	Airline	UK	Data analytics applications	Review
Yu Shi et al.	2021	To provide a comprehensive review of the applications of data science techniques and methodologies in productivity	Usage of data science techniques in productivity has been growing since 2005	General	UK		Review
ShemilaAbbasi et. al.	2022	To spotlight the necessity of electronic health record system and to draw the	Automated information systems (AIMS) records the events taking place during the preoperative phase (clinical procedures,	Healthcare	Pakistan	Electronic Health Record Systems (EHRS)	Conceptual

		attention of the concerned authorities to plan provision of it in operating rooms.	physiologic parameter changes, and medication administration) and stored in a relational database management system (RDBMS) which stores and provides inferential data.			Anesthesia Information Management System (AIMS). RDBMS	
Kataria et al.	2022	To illustrate macro factors, advantages, AI (Artificial Intelligence) applications and data analytics in the aspect of patient care for human data science.	The key macro factors are: Biological Process, mental health, social interactions and cultural forces responsible for data diversity, personal genetic signature, and day-to-day impact on variability. Application of AI and data analytics: study design assessment, monitoring and progress, deconstruct and illustrate data at trial design stage. Data mining, patient matching and Monitoring, analysis of various parameters and data sharing automation at trial initiation stage. Finally, data cleaning, analysis and visualization at the stage of completion of clinical trials.	Healthcare	India	Artificial Intelligence (AI) and Data Analytics Data Mining	Review
Aaron Morelos-Gomez et al.	2022	To study the characteristics, synthesis & application of carbon materials using data science tools & techniques	*Materials synthesis in conjunction with data science has resulted in optimal growth conditions with desired properties and processing techniques * Regarding characterization, molecular structures can be reconstructed using microscopy images * applications of carbon materials, data science has enabled prediction of the water treatment efficiency, classification of electronic signals, prediction of the biological activity, and virus classification	Environmental	Japan	Regressions, and Microscopic images	Review
Barbosa-Silva et al	2022	Analyzed the genes of the WNT pathway and seven cross-linked pathways that may explain the differences in aggressiveness among cancer types	They found GRB2, CTNNB1, SKP1, CSNK2A1, PRKDC, HDAC1, YWHAZ, YWHAB, and PSMD2. Except for PSMD2, the RFC analysis showed a different list, which was CAD, PSMD14, APH1A, PSMD2, SHC1, TMEFF2, PSMD11,	Healthcare	Switzerland	R Analytic Flow version 4	Empirical

			H2AFZ, PSMB5, and NOTCH1				
Edward Hoornstra et al	2022	To quantify the effectiveness of propofol infusion when administered either via total intravenous administration (TIVA) or combined intravenous anesthesia (CIVA) with inhalational agents on PONV.	Propofol infusion has a naive effect on PONV with a lift of -41% (P < .001) when using TIVA and -17% (P < .001) when using CIVA.	Healthcare	USA	Logistic regression models	Empirical
Pérez-Ortega et al.	2022	To analysis clusters related to mortality rate from COVID-19 at the municipal level in Mexico from the perspective of Data Science.	Two key indicators related to mortality from COVID-19 at the municipal level were identified: one is population density and the other is percentage of population in poverty. clusters with high values of mortality rate had high values of population density and low poverty levels	Healthcare	USA	Batch Foundation Methodology for Data Science (FMDS)	Empirical
Rautenbach et al.	2022	To identify barriers to implementation of DS to support sustainable business operations in SMEs.	Data quality concerns, Insufficient infrastructure, Financial constraints, Data privacy and security concerns, Social challenges, Access to software, Lack of skills	SMEs	Republic of South Africa	Content Analysis	Review
Lund, Brady and Ting Wang	2022	To review literature pertaining to the development of data science as a discipline, current issues with data bias and ethics	Information science researchers have already contributed to a humanistic approach to data ethics within the literature and an emphasis on data science within information schools all but ensures that this literature will continue to grow in coming decades	Education	USA		Review
Yeonji Jung et al.	2022	To describe a theory-informed application of data science methods to analyze the quality of reflections made in a health professions education program over time	A dramatic increase from No to Shallow reflections from the start to end of year one (20% → 66%), but only a limited gradual rise in Deep reflections across all four years (2% → 26%). The presence of all six reflection elements increased over time, but inclusion of Feelings and Analysis remained relatively low even at the end of year four (found in 44% and 60%	Education	USA	Machine learning models	Empirical

			of reflections respectively).				
Swapnil Morande	2022	To explore parameters related to an individual's cognitive interactions to manage stress.	A significant impact of age, gender and the state of health on stress. Variables associated with lifestyle brainwaves and therapeutic intervention positively influence stress levels.	Healthcare	Italy	Machine Learning model (BigML)	Empirical
Deepti Chopra, and Praveen Arora	2022	To discuss the scope of Swarm Intelligence in data sciences	In Data Science, Swarm Intelligence is mainly used for optimization of Data or tuning the parameter that may involve certain statistics or machine learning technique	General	India	Swarm Intelligence (meta-heuristics algorithms)	Conceptual
Jalajakshi and Myna	2022	To discuss the importance and contribution of statistics to Data science.	Statistics is proved to be an important discipline in regulating the work analyzed in the field of Data Science.	Education	India	Linear Regression	Conceptual
Cussat-Blanc et al.	2023	To discuss some necessary changes in the health studies curriculum that could help practitioners to interpret decisions the made by a machine.	AI/ML play a major role in precision medicine. The MDS should not be considered alone but with the coordination of mathematicians and computer scientists within a new department. the MDS department should manage the local, regional, national and even international integration of data to develop ML models	Healthcare	Switzerland	Machine Learning (ML) and Artificial Intelligence (AI)	Conceptual
Duncan et al.	2023	To evaluate the presence of IP among data science students and to evaluate several variables linked to IP simultaneously in a single study evaluating data science.	Most students in the sample showed moderate and frequent levels of IP. Gender identification was positively related to IP for both males and females.	Education	USA	Multivariate Analysis of Variance (MANOVA)	Empirical
M. Baillie et al.	2023	To discuss obstacles with the aim of opening a dialogue on good data science practice in the context of drug development.	Data Science is often thought of as a synonym for machine learning or predictive analytics. But this is a limited view. Data science is only the current label that is a trend toward the integration of computational and statistical practices.	Drug Industry	Switzerland	---	Conceptual
Capobianco and Dominietto	2023	To emphasize medical imaging and radiomics as the leading contextual	Radiomics will be very useful in dealing with specific tumors, for instance, brain or head & neck, given the possibility to produce	Healthcare	Switzerland	Artificial Intelligence (AI) and Machine	Review

		frameworks to measure the impacts of Artificial Intelligence (AI) and Machine Learning (ML) developments.	consistent acquisition protocols that facilitate reproducible results. Radiomics contribute more systematically to the process of selecting optimal treatments.			Learning (ML)	
Glushko, J. Robert	2023	To discusses ways to make a data science project fail.	No clear goals or problem statements, Missing skills, Problems with data, Over-reliance with technology, Poor deployment planning, Poor maintenance planning are the seven ways of Unrealistic expectations.	Education	USA		Conceptual
Marchionini, Gary	2023	To compare information and data science with respect to knowledge, area of study and practice.	In data science, context can be defined, often through quantization or dimensionality reduction; minimal metadata while in information science, context is social and probabilistic at best; significant attention to both metadata for system use and human consumption	Education	USA		Conceptual
Meulemeester and Martens	2023	To estimate the contribution of the increasingly popular "common" data science to the global carbon footprint.	"common" data science consumes 2.57 more power than regular computer usage, but less than some common everyday power-consuming tasks such as lighting or heating,	Power Consumption	Belgium	AI and Carbon emission	Empirical
NaiyarIqbala, and Pradeep Kumarb	2023	To explore the role of data science and soft computing approaches in bioinformatics disciplines with various associated applications	Simple soft computing techniques involves fuzzy logic, artificial neural networks, support vector machines and evolutionary computation techniques in bioinformatics disciplines such as chemistry, engineering, mathematics and computer sciences.	Education	India	Fuzzy Logic, Artificial neural network, support vector machine	Conceptual

4.1 Data Science Definitions/Meanings

Table 2 shows the definitions of data sciences. Data science is the intersection of various fields (computer science, business engineering, statistics, data mining, etc.) and a machine learning technique to solve problems, develop insights, identify correlations, causal relationships, patterns and anomalies, classify and predict events, and infer probabilities, interest and sentiments.

Table 2: Data Science Definitions/Meanings

Definitions/Meanings	Researcher (s)	Year
The analysis of data to solve problems and develop insights	Saltz et al.	2018
A process of multi-stage knowledge discovery in which useful knowledge is extracted from a raw, often impure collection in a specific context	Turkay et al.	2018
Provides principles, methodology, and guidance for data analysis for: (1) an instrument (visualization, data collection, or research tools), (2) value (commercial or scientific), or (3) knowledge (hidden objective practical useful interdependencies)	Donghui& Davis,	2019
'Data Science' is a cyclical process of capturing the business needs and acquiring the relevant data, storage, security, privacy, preparation, pre-processing, analytics; generating and communicating insights; and finally, actuating these actionable insights	Jägare	2019

4.2 Contexts and Country-Wise Publications

Research publications dates ranged from 2018 to 2023 with the number of studies conducted in different countries. Table 3 illustrates area of study published across varying countries. Researchers observed that research articles were published across 20 countries with the majority of United States of America (16) followed by United Kingdom and Switzerland (5) each. The rest of the countries published one or two research work. It is also revealed that data science research work has been mostly conducted in the education context followed by healthcare context, general, SMEs and airlines.

Table 3: Contexts and Country-Wise Publications

Country	Contexts	Numbers	Percentages
India	Healthcare	01	2.5
	General	02	5
	Education	01	2.5
USA	Healthcare	01	2.5
	Education	15	37.5
Japan	Environmental	01	2.5
United Kingdom	Healthcare	02	5
	Education	01	2.5
	Airlines	01	2.5
	General	01	2.5
Saudi Arabia	Healthcare	01	2.5
Switzerland	Education	01	2.5
	Healthcare	04	10
South Africa	SMEs	01	2.5
Italy	Healthcare	01	2.5
Pakistan	Healthcare	01	2.5
China	Education	01	2.5
Spain	General	01	2.5
Russia	General	01	2.5
Belgium	Power Consumption	01	2.5
Argentina	Healthcare	01	2.5
	Total	**40**	**100**

4.3 Research Areas/Contexts Publications

Table 4 depicts that education (19) is the most frequently studied research area indicating that more than half of the studies are focused on teaching issues in the business, medicine and economic sectors. The other highest area of research conducted in the healthcare (12) sector followed by general studies, airlines, power consumptions, environmental and SMEs.

Table4: Research Ares Publications

Contexts	Numbers	Percentages
Healthcare	12	30
Education	19	47.5
Environmental	01	2.5
General	05	12.5
Airlines	01	2.5
SMEs	01	2.5
Power Consumption	01	2.5
Total	**40**	**100**

4.4 Year-Wise and Types of study

The objective of the research was to know what type of study was conducted during the 2018 to 2023. Table 5 reveals that most of the studies were empirical (17) followed by conceptual (13) and Reviews (12). Number of studies gradually increases from 2018 till 2021. But, gradually decreased from 2021 till 2023. However, the maximum numbers of studies were 14 in the year 2021.

Table 5: Year-Wise and Types of study

Years	Types of study			
	Review	Empirical	Conceptual	Total
2018	0	1	1	2
2019	1	0	1	2
2020	1	3	2	6
2021	5	7	2	14
2022	4	5	3	12
2023	1	1	4	6
Total	**12**	**17**	**13**	**40**

4.5 Most Relevant Sources

The researchers decided to know the most relevant sources for publishing research work of data sciences. Table 6 reveals the most relevant sources of pushing research work.

Table 6: Most Relevant Sources

Most Relevant Sources	
BMC Musculoskeletal Disorders	The South African Journal of Industrial Engineering
Biosciences	International Journal of Environmental Research & Public Health
BMC Bioinformatics	International Journal of Information Management
John Wiley and Sons	Statistics in Biopharmaceutical Research
Anesthesia, Pain and Intensive Care	Journal of data and information science
Korean Journal Anesthesiology	Advances in Health Sciences Education
Communication of ACM	Journal of Clinical Medicine
Academy Management Journal	Journal of surgical oncology

Journal of Documentation	Transport Reviews
Procedia Computer Science	Neural Computing and Applications
Expert System	Global Transitions Proceedings
Big Data	Data and Information Management
TEM Journal	Informatics and Systems
Cancers	Journal of Statistics Education
AANA Journal	Advances in Health Science Education
Mathematics	Information System Education Journal
Briefing in Bioinformatics	International Journal of Recent Technology and Engineering
Teaching Statistics	Advance Theory and Simulations

4.6 Data Science Tools/Techniques

The researchers also identified the tools and techniques of data sciences that are studied by the research scholars in various fields (business, education, healthcare, airlines, environment, and power consumption). Table 7 shows the tools and techniques of data sciences.

Table 7: Data Science Tools/Techniques

Tools/Techniques		
Artificial Intelligences	pyKVFinder	Deep Neural Network (BS-DNN)
Machine Learning's	JSON	Computer base Algorithms
ANNOVA	Artificial Neural Network	Big Data
Regressions	Support Vector Machines	SQL
Cross Tabulations	Clinical Imaging	CODAP
Bayesian Theory	Phylogenetic analysis	Data Mining
R and R Markdown	IPO (Improved Pigeon Optimization Algorithms)	R Analytic Flow
CRISP-DM data cycle	R	Content Analysis
Python	Data Analytics	Swarm Intelligence
SAS	RDBMS	MANOVA

5. DISCUSSIONS AND CONCLUSIONS

The main purpose of this study was to review past research studies during 2018-2023 in the context of data science in the various fields. The other aim was to analyze the definitions, country-wise publications in different context, research area of studies, year-wise studies, different types of studies, tools and techniques and the most relevant sources of publishing research work in the context of data science.

The notion data science has defined as the intersection of computer science, business, statistics, engineering, data mining, machine learning, six sigma, automation, and operation research and domain expertise (Vincent Granville (2014) that support and guide the principled extraction of information and knowledge from the data (Provost and Fawcett, 2013). In addition, it is defined as a statistical and machine learning techniques on big multi-structured data in a distributed computing environment to identify correlations and causal relationships, classify and predict events, identify patterns and anomalies, and infer probabilities, interest and sentiment (Manripu Das et al., 2015). For them, data science combines expertise across software development, data management and statistics. The main purpose of data science analyze is to solve problem (s) and develop insights (Slatz et al., 2018).

In addition, the researchers identified the publications of research studies in various countries. Research studies of data science were published across 20 countries with the majority of United States of America followed by United Kingdom, Switzerland, India, Saudi Arabia, Italy, Pakistan, China, Spain, Russia, Belgium, Argentina and Japan. Majority of research studies conducted in the context of Education followed by HealthCare, General, Environmental, Airlines, SMEs and Power consumption only. The researchers also identified the different types of research work. Majority of the studies were found empirical followed by conceptual and reviews and most of studies conducted in the year 2021 and 2022.

Furthermore, the most relevant sources of studies found in the journals of medical, Health science, Computer science, and business. The tools of data science used in the studies are as follows: Artificial Intelligence, Deep Neural Network (BS-DNN), Machine Learning's, Artificial Neural Network, Support Vector Machines, Improved Pigeon Optimization Algorithms (IPO), R Analytic Flow, CODAP, SQL, CRISP-DM data cycle, data mining, python, SAS, Phylogenetic analysis and Swarm Intelligence.

REFERENCES

1) Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. International Journal of Social Research Methodology, 8(1), 19–32. https://doi.org/10.1080/1364557032000119616
2) Aaron Morelos-Gomez, Mauricio Terrones, and Morinobu Endo. (2022). Data Science Applied to Carbon Materials: Synthesis, Characterization, and Applications, Advance Theory and Simulations, Volume 5 and Issue 2.
3) Barbosa-Silva, A.; Magalhães, M.; da Silva, G.F.; da Silva, F.A.B.; Carneiro, F.R.G.; Carels, N. A Data Science Approach for the Identification of Molecular Signatures of Aggressive Cancers. Cancers 2022, 14, 2325. https://doi.org/ 10.3390/cancers14092325
4) C. Turkay, N. Pezzotti, C. Binnig, H. Strobelt, B. Hammer, D. A. Keim, J.-D. Fekete, T. Palpanas, Y. Wang, F. Rusu, "Progressive data science: Potential and challenges," 2018. Retrieved from https://hal.inria.fr/hal-01961871/document
5) C. Turkay, N. Pezzotti, C. Binnig, H. Strobelt, B. Hammer, D. A. Keim, J.-D. Fekete, T. Palpanas, Y. Wang, F. Rusu, "Progressive data science: Potential and challenges," 2018. Retrieved from https://hal.inria.fr/hal-01961871/document
6) Capobianco, E.; Dominietto, M. Translating Data Science Results into Precision Oncology Decisions: A Mini Review. J. Clin. Med. 2023, 12, 438. https://doi.org/10.3390/ jcm12020438.
7) Cussat-Blanc, S.; Castets-Renard, C.; Monsarrat, P. Doctors in Medical Data Sciences: A New Curriculum, International Journal of Environmental Research and Public Health 2023, 20, 675. https:// doi.org/10.3390/ijerph2001067
8) D.M. Dedge Parks, "Defining Data Science and Data Scientist," Graduate Theses and Dissertations, 2017. Retrieved from http://scholarcommons.usf.edu/etd/7014
9) Deepti Chopra and Praveen Arora. (2022). Swarm Intelligence in Data Science: Challenges, Opportunities and Applications, in 4th International Conference on Innovative Data Communication Technology and application, Procedia Computer Science 215 (2022) 104–111.
10) Dhar, V. (2013), "Data Science and Prediction", Communications of the ACM, Vol. 56 No. 12, pp. 64-73

11) DongwooChae, Data Science and Machine Learning in Anesthesiology, Korean Journal of Anesthesiology 2020, No.73, Issue (4), pp. 285-295.

12) Duncan, L.; Taasoobshirazi, G.; Vaudreuil, A.; Kota, J.S.; Sneha, S. An Evaluation of Impostor Phenomenon in Data Science Students. Int. J. Environ. Res. Public Health 2023, 20, 4115. https:// doi.org/10.3390/ijerph20054115.

13) Edward Hoornstra, NishantVelagapudi, John T. Bryant, Douglas Roberts, LakshmanaBehara, RavithejaVenigandla, Prakash Mani, Bala G. Nair, Application of Data Science to Quantify the Effect of Propofol Infusion on Postoperative Nausea and Vomiting, AANA Journal + August 2022 + Vol. 90 No. 4, pp. 263-270.

14) Emery, C. Nathan, Erika Crispo, Sarah R Supp, Kaitlin J Farrell, Andrew J Kerkhoff, Ellen K Bledsoe, Kelly L O'Donnell, Andrew C McCall, Matthew E Aiello-Lammens, (2021). Data Science in Undergraduate Life Science Education: A Need for Instructor Skills Training,*BioScience*, Volume 71, Issue 12, December 2021, Pages 1274–1287, https://doi.org/10.1093/biosci/biab107.

15) George, G., Osinga, E.C., Lavie, D. and Scott, B.A. (2016), "Big Data and Data Science Methods for Management Research", Academy Management Journal, Vol. 59 No. 5, pp. 1493-1507

16) Glushko, J. Robert. (2023). Seven Ways to Make a Data Science Fail, Data and Information Management 7 (2023) 100029, pp. 1-5.

17) Jägare U (2019) Data science strategy for dummies. John Wiley and Sons, Incorporated

18) Jägare U. (2019). Data Science Strategy for Dummies. John Wiley and Sons, Incorporated

19) João Victor da Silva Guerra1, HelderVerasRibeiro-Filho, Gabriel Ernesto Jara, Leandro Oliveira Bortot, José Geraldo de Carvalho Pereira and Paulo Sérgio Lopes-de-Oliveira1 (2021). pyKVFinder: An Efficient and Integral Python Package for Bimolecular Cavity Detection And Characterization In Data Science, BMC Bioinformatics, 22:607

20) Kataria, Suchitra, Ravindran, Vinod (2022).Musculoskeletal Care - at the Confluence of Data Science, Sensors, Engineering, and Computation, BMC Musculoskeletal Disorders (2022) 23:169.

21) Kirby McMaster, Brian Rague, Stuart L. Wolthuis, and Samuel Sambasivam. (2018). Information Systems Education Journal (ISEDJ), 16 (1) ISSN: 1545-679X.

22) Kumar Das, Jayanta; Giuseppe Tradigo, PierangeloVeltri and Pietro H Guzzi and Swarup Roy (2021). Data Science in Unveiling COVID-19 Pathogenesis and Diagnosis: Evolutionary Origin to Drug Repurposing, *Briefings in Bioinformatics*, Volume 22, Issue 2, March 2021, Pages 855–872, https://doi.org/10.1093/bib/bbaa420

23) Larson, D. and Chang, V. (2016), "A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science", International Journal of Information Management, Vol. 36, pp. 700-710.

24) Leonardo Carvalho, Alice Sternberg, Leandro Maia Gonçalves, Ana Beatriz Cruz, Jorge A. Soares , Diego Brandão , Diego Carvalho and Eduardo Ogasawara. (2021). On the Relevance of Data Science for flight Delay Research: A Systematic Review, TRANSPORT REVIEWS 2021, VOL. 41, and NO. 4, 499–528 https://doi.org/10.1080/01441647.2020.1861123.

25) Lin Wang, (2018). Twinning Data Science with Information Science in Schools of Library and Information Science, Journal of Documentation Vol. 74 No. 6, pp. 1244-1256.

26) Lund, B., & Wang, T. (2022). What does Information Science Offer for Data Science Research? A Review of Data and Information Ethics Literature. Journal of data and information science, 7(4), 16–38. https://doi.org/10.2478/ jdis-2022-0018.

27) Marchionini, Gary. (2023). Information and Data Sciences: Context, Unit of Analysis, Meaning and Human Impact, Data and Information Management 7 (2023) 100031, pp. 1-5.

28) Maria Jose´ Sousa, Pere Mercade´ Mele´ , Anto´nio Miguel Pesqueira, A´ lvaro Rocha, Miguel Sousa, Salma Noor, (2021) Data Science Strategies Leading to the Development of Data Scientists' Skills In Organizations, Neural Computing & Applications, 2021, 33 (1).

29) Martin G. Tolsgaard1, Christy K. Boscardin, Yoon Soo Park, Monica M. Cuddy, Stefanie S. Sebok-Syer. (2020). The Role of Data Science and Machine Learning in Health Professions Education: Practical Applications, Theoretical Contributions, And Epistemic Beliefs, Advances in Health Sciences Education (2020) 25:1057–1086 https://doi.org/10.1007/s10459-020-10009-8.

30) Meulemeester, Bjoerg and Martens, David. (2023). How Sustainable is "Common" Data Science in Terms of Power Consumption, Sustainable Computing: Informatics and Systems 38 (2023) 100864, pp. 1-9.

31) MujthabaG.M ,Abdalla Al Ameen, ManjurKolhar, Mohammed Rahmath. (2020). International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-6.

32) Naiyar Iqbal and Pradeep Kumar. (2023). From Data Science to Bioscience: Emerging Era of Bioinformatics Applications, Tools and Challenges, In International Conference on Machine Learning and Data Engineering, Procedia Computer Science 218 (2023) 1516–1528.

33) P. J. Diggle, "Statistics: A Data Science for the 21st Century," Royal Statistical Society, Series A (Statistics in Society), vol.178, part 4, pp. 793-813, 2015. Retrieved from https://www.researchgate.net/publication/323393464_Data_Science_Literature_Review_State_of _Art

34) Pérez-Ortega,J.; Almanza-Ortega,N.N.;Torres-Poveda, K.;Martínez-González,G.; Zavala-Díaz,J.C.;Pazos-Rangel,R. Application of Data Science for Cluster Analysis of COVID-19 Mortality According to Socio demographic Factors at Municipal Level in Mexico. Mathematics 2022, 10, 2167. https:// doi.org/10.3390/math10132167

35) Provost, F. and Fawcett, T. (2013), "Data Science and its Relationship to Big Data and Data-Driven Decision Making", Big Data, Vol. 1 No. 1, pp. 51-59.

36) Rautenbach, S., de Kock, I., &Grobler, J. (2022). Data Science for Small and Medium-Sized Enterprises: A Structured Literature Review. *The South African Journal of Industrial Engineering*, *33*(3), 83–95. https://doi.org/10.7166/33-3-2797

37) S. Shum, R. Baker, J. Behrens, M. Hawksey, N. Jeffery, R. Pea, "Educational Data Scientists: A Scarce Breed," 2013. Retrieved from http://simon.buckinghamshum.net/wpcontent/uploads/2013/03/LAK13PanelEducDataScientists.

38) SanketMantri. (2018). Data Science: Literature Review & State of Art

39) Sergey V. Novikov. (2020). Data Science and Big Data Technologies Role in the Digital Economy,
TEM Journal. Volume 9, Issue 2, Pages 756-762, ISSN 2217-8309, DOI: 10.18421/TEM92-44, May 2020.

40) ShemilaAbbasi, Usama Ahmed, FauziaAnis Khan, Data Science and its Application in Anesthesiology, Anesthesia, Pain & Intensive Care, Vol 26 (1); February 2022, pp.1-2.

41) Song, I. and Zhu, Y.J. (2016), "Big Data and Data Science: What Should we Teach?" Expert Systems, Vol. 33 No. 4, pp. 364-373.

42) VarlamKutateladze, and Ekaterina Seregina, "Fast and Efficient Data Science Techniques for COVID-19 Group Testing", Department of Economics, University of California, Riverside, CA 92521, USA.

43) Y. Donghui, E. Davis, "A First Course in Data Science," 2019. Retrieved from https://arxiv.org/pdf/1905.03121.pdf

44) Yeonji Jung, Alyssa Friend Wise, Kenneth L. Allen. (2022). Using Theory-Informed Data Science Methods To Trace The Quality Of Dental Student Reflections Over Time, Advances in Health Sciences Education (2022) 27:23–48 https://doi.org/10.1007/s10459-021-10067-6.