

ПРИЛОЖЕНИЕ НА АНАЛИЗА НА ДАННИ ЗА МОДЕЛИРАНЕ НА ОБУЧЕНИЕТО

Нели Христова
Иван Попчев
Даниела Орозова
Бургаски свободен университет

APPLICATION OF DATA ANALYSIS FOR MODELING OF EDUCATION

Nely Hristova
Ivan Popchev
Daniela Orozova
Burgas Free University

Abstract: *The purpose of this research is to examine how the data science, big data, data analysis, data science in education, data science in mathematics, and data science in mathematical education are presented in the Computer Database (Gale), Applied Science & Technology Source (EBSCO) and www.sciencedirect.com databases. We use the following key words: data science, big data, data analysis, data science in education, data science in mathematics, and data science in mathematical education. The examined period is from 1996 to 2018.*

Key words: *data science, big data, data analysis, data science in education, data science in mathematics, data science in mathematical education*

I. Мотивация на изследването

Всички сме потопени в данни. Данните се предават от всяка дейност, която се осъществява в дигиталния и физическия свят. Все повече те могат да се използват за разбиране и подобряване на бизнеса, инвестициите, здравеопазването, начина на живот. Науката за данните (Data Science) помага да се предвиди броя на учениците в гимназиите, да се прогнозира силата и пътя на ураганите, да се опознават и предвиждат различни процеси. Има толкова много примери как може да се използва науката за данните за организиране и подобряване на ежедневието. Значението и въздействието на анализите на данните вече са по-големи от всякога. Data Science представя подходи, на базата на които се създават модели, откриват се взаимоотношения и се дава смисъл на данните.

II. Цел на изследването

Целта на настоящото изследване е да се очертаят тенденциите в анализа на данните от последните години, което от своя страна да послужи като основа за бъдещи по-задълбочени и по-тясно специализирани анализи. Количествена мярка за обособяване на тези тенденции е броят на публикациите, представени в *Computer Database*

(Gale) [9], *Applied Science & Technology Source (EBSCO)* [10] и *www.sciencedirect.com* [8] от началото на 1996 година до началото на 2019 година. Ключовите думи, по които е направено проучването са: **data science, big data, data analysis, data science in education, data science in mathematics, и data science in mathematical education.**

Направеното проучване показва, че интересът към науката за данните расте, че тази наука е особено важна за образованието и в частност в обучението по математика. Данните са обособени в таблици и графики за по-добро визуално представяне.

III. Изследване

Нека да проучим какво е състоянието в научната база от данни *Computer Database (Gale)* относно: data science (наука за данните), big data (големи данни), data analysis (анализ на данните), data science in education (наука за данните в образованието), data science in mathematics (наука за данните в математиката) и data science in mathematical education (наука за данните в обучението по математика).

Резултатите са отразени в таблица 1.

Таблица 1

	1996 година	2000 година	2018 година
Data Science	Списания 45	Списания 73	Списания 175
	Академични журналы 34	Академични журналы 55	Академични журналы 258
	Новини 31	Новини 29	Новини 2053
Big Data	Списания 124	Списания 228	Списания 694
	Академични журналы 19	Академични журналы 23	Академични журналы 545
	Новини 44	Новини 48	Новини 2772
Data Analysis	Списания 671	Списания 1123	Списания 382
	Академични журналы 207	Академични журналы 465	Академични журналы 724
	Новини 224	Новини 300	Новини 4716
Data Science in Education	Списания 1	Списания 0	Списания 4
	Академични журналы 1	Академични журналы 0	Академични журналы 0
	Новини 0	Новини 0	Новини 1
Data Science in Mathematics	Списания 0	Списания 0	Списания 0
	Академични журналы 0	Академични журналы 0	Академични журналы 1
	Новини 0	Новини 0	Новини 6
Data Science in Mathematical Education	Списания 0	Списания 0	Списания 0
	Академични журналы 0	Академични журналы 0	Академични журналы 0
	Новини 0	Новини 0	Новини 0

От данните в таблицата може да се направи извода, че броят на публикациите в областите: data science (наука за данните), big data (голям обем от данни) и data analysis (анализ на данните) расте от 1996 година до 2018 година включително. Но в

областите data science in education (наука за данните в образованието), data science in mathematics (наука за данните в математиката) и data science in mathematical education (наука за данните за обучението по математика) няма почти никакви публикации.

Нека да проучим какво е състоянието в научната база от данни *Applied Science & Technology (EBSCO)* относно: data science (наука за данните), big data (голям обем от данни), data analysis (анализ на данните), data science in education (наука за данните в образованието), data science in mathematics (наука за данните в математиката) и data science in mathematical education (наука за данните за обучението по математика). Резултатите са отразени в таблица 2.

Таблица 2

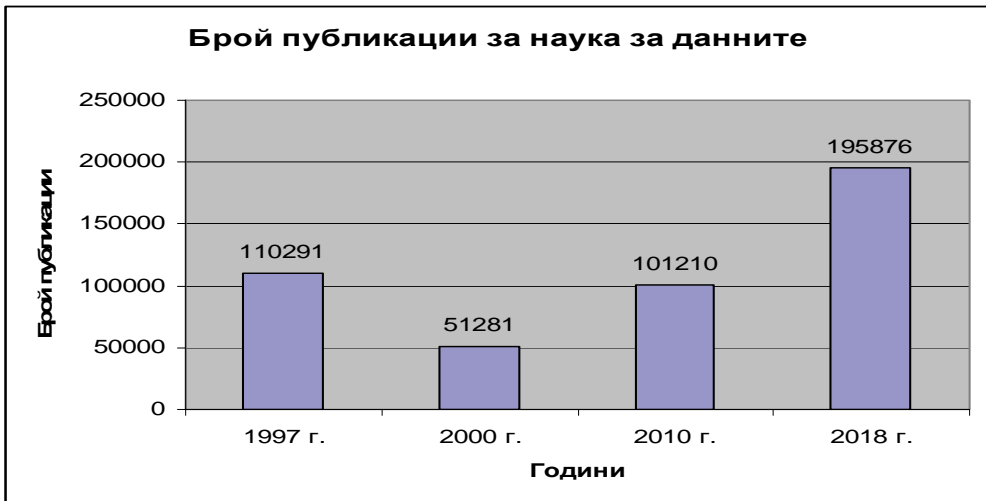
	1996 година	2000 година	2018 година
Data Science	Списания 45	Списания 72	Списания 92
	Академични журнари 150	Академични журнари 226	Академични журнари 2060
	Ревю на книги 1		Ревю на книги 13
			Материали, представени на конференции 22
Big Data	Списания 5	Списания 3	Списания 105
			Академични журнари 2607
			Ревю на книги 13
			Материали, представени на конференции 45
Data Analysis	Списания 115	Списания 75	Списания 168
	Академични журнари 605	Академични журнари 866	Академични журнари 5813
	Ревю на книги 21	Книги 2	Ревю на книги 10
			Материали, представени на конференции 56
Data Science in Education	Няма резултати	Списания 1	Списания 1
			Академични журнари 9
Data Science in Mathematics	Академични журнари 1	Академични журнари 1	Академични журнари 32
Data Science in Mathematical Education	Няма резултати	Няма резултати	Няма резултати

От таблицата се вижда, че броят на публикациите в областите: data science (наука за данните), big data (голям обем от данни) и data analysis (анализи на данните) расте от 1996 година до 2018 година включително. Но в областите: data science in education (наука за данните в образованието) и data science in mathematics (наука за данните в математиката) публикациите са съвсем малко, като за data science in mathematical education (наука за данните за обучението по математика) няма публикации.

Следващата научна база, разгледана в изследването е Science Direct. В www.sciencedirect.com търсим броя на публикациите от 1997 година до края на 2018 година относно: data science (наука за данните), big data (голям обем от данни), data analysis (анализ на данните), data science in education (наука за данните в образованието), data science in mathematics (наука за данните в математиката) и data science in mathematical education (наука за данните за обучението по математика). Тук най-ранната година е 1997 година, поради това нашата база за сравнение е 1997 година.

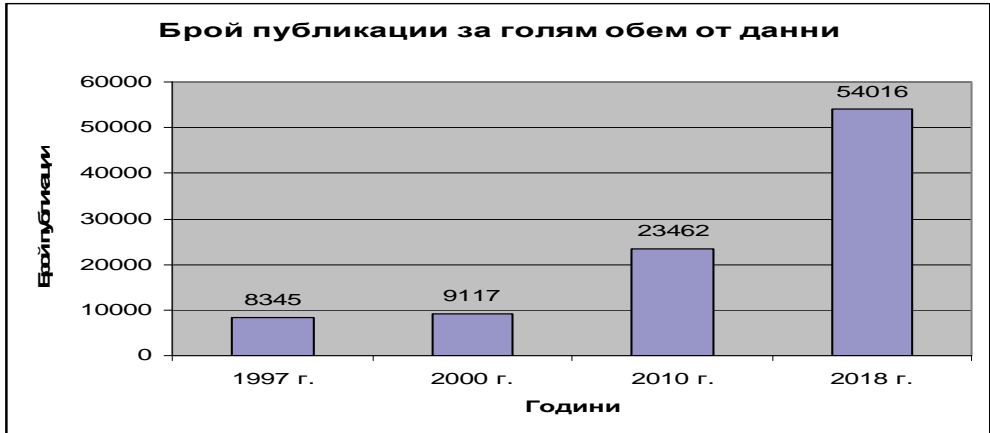
Резултатите за броя на публикациите в www.sciencedirect.com, отнасящи се за термините: *наука за данните (data science)* са обобщени с графика 1. Ясно се вижда, че броят публикации намалява от 1997 г. до 2000 г., когато е абсолютният минимум. След 2000 г. обаче този брой расте много бързо и се проявява в устойчива тенденция на разтеж.

Графика 1



В www.sciencedirect.com резултатите за публикации, отнасящи се за термина *големи данни (big data)* са обобщени в графика 2. Техният брой стремително расте от 1997 г. до 2018 г. Тази устойчива тенденция показва, че анализа на големи данни е полезен за разкриване на тенденции и използването им за предвиждане на събития.

Графика 2



В www.sciencedirect.com резултатите за публикации, отнасящи се до термина: *анализ на данните (data analysis)* са обобщени в графика 3. Виждаме същата тенденция както при *големите данни (big data)*. Броят публикации за *анализ на данните (data analysis)* силно расте от 1997 до 2018 г. включително.

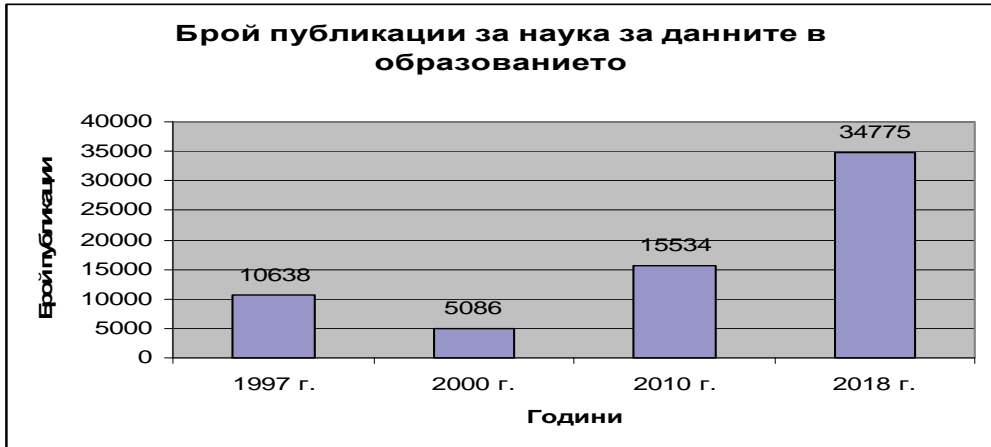
Графика 3



Науката за данните се използва и в *областта на образованието (data science in education)*.

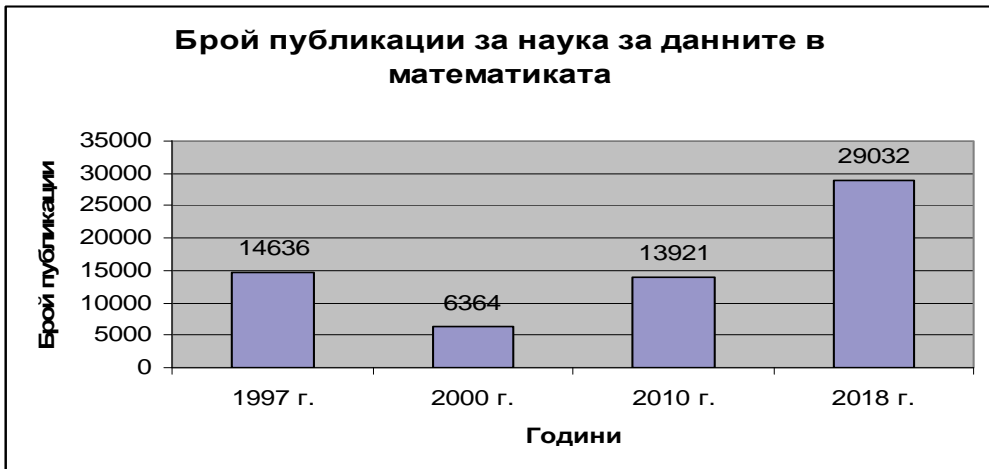
Резултатите за броя на публикации в сферата *науката на данните в образованието* са обобщени в графика 4. От 1997 г. до 2000 г. има известен спад в броя публикации в тази област, но след това има стабилна тенденция на растеж. Сферата на образованието прилага анализи на данни, за да планира по-добре обучаващите процеси, да оптимизира учебните планове, да предвиди броя на учениците и преподавателите, да разпредели средствата за образование и редица други задачи. В сферата на образованието анализа на данните се използва за много и различни приложения.

Графика 4



Науката за данните в математиката също е много актуална тематика, както се вижда от данните за броя от публикациите, свързани с термина *data science in mathematics*, обобщени в графика 5. Отново от 1997 г. до 2000 г. има известен спад в броя публикации в областта, но след това има стабилна тенденция на растеж, която продължава и до сега. В областта на математиката ясно се очертава тенденцията за анализ на данните и получаване на обобщена информация от тях с различни приложения.

Графика 5



Следва представяне на данните относно броя на публикациите, свързани с науката за *данните в обучението по математиката (data science in mathematics)*. Тези данни са обобщени в диаграмата от графика 6. Отново от 1997 г. до 2000 г. има известен спад в броя публикации в областта, но след това има стабилна тенденция на растеж, която продължава и до сега. Това показва, че и в областта на обучението по математика науката за данните намира все по-широко приложение.

Графика 6



Следва кратко разглеждане на някои приложения от статиите, публикувани в *Procedia Computer Science* за науката за данните в обучението по математика.

- В статията *Primary Education Evaluation in Brazil using Big Data and Cluster Analysis* от Thiago Graca Ramosa, Jean Cristian Ferreira Machadob, Bruna Principe и Vieira Cordeiroc се синтезират резултати за качеството на образованието по отношение на стандартизираното и средното представяне на учениците по португалски език и математика. На всеки две години правителството оценява обучението в училищата. Изследването използва данните от четири оценявания, Big Data, Warehouse and cluster analysis, за да анализира обучението в страната и да се дадат предложения за възможни подобрения и промени. Това е статия със свободен достъп [1].
- В статията *Pattern Analysis in the Study of Science, Education and Innovative Activity in Russian Regions* от F. Aleskerov, L. Egorova и др. се разглежда метода на анализ на шаблони (pattern analysis) и резултатите от неговото приложение за възможностите в областта на науката и образованието в регионите на Руската Федерация [2].
- В статията *A Study on the Application of Big Data to the Korean College Educational System* от Yeon Hee Kim и Jin-Ho Ahn се представя как метода „learning analysis“ се използва, за да се анализира голям обем от данни с цел подобряване на обучението в Корея [3].
- В статията *Learning Curve Analysis Using Intensive Longitudinal and Cluster Correlated Data* с автори Xiao Zhonga, Zeyi Suna, Haoyi Xionga, Neil Heffernanb, Md. Monirul Islama от Missouri University of Science and Technology, Rolla, USA се представят методите GLMs и MANOVA, които са приложени, за да се оцени онлайн обучението чрез Wald тест. Целта е да се оцени ефекта от обучението на студентите, които са изучавали предмета Алгебра в продължение на три месеца [4].

Подход за наблюдение и събиране на данни за студентите и тяхната дейност по време на дистанционното им обучение и анализ на натрупаните данни дава добри резултати при локализацията на студентите в риск. Това е многоетапна процедура, представена в [5], която е пряко свързана с анализа на големите данни в пространството за електронно обучение.

IV. Заключение

Науката за данните заема съществено място в съвременния живот и нейното разбиране и използване неминуемо ще нараства. Бизнесът и образованието виждат тенденциите и ползата от анализите на големите данни. Проведеното изследване налага този анализ да бъде продължен във времето и да бъде разширен, като обхване и редица други световни научни бази от данни.

Литература:

- [1] Ramos T., Machado J., etc., Primary Education Evaluation in Brazil Using Big Data and Cluster Analysis, *Procedia Computer Science*, Volume 55, 2015, Pages 1031-1039
- [2] Alexandrov F, Egorova L, etc., Pattern Analysis in the Study of Science, Education and Innovative Activity in Russian Regions, *Procedia Computer Science*, Volume 17, 2014, Pages 678-694
- [3] Yeon Hee Kim, Jin-Ho Ahn, A Study on the Application of Big Data to the Korean College Education System
Procedia Computer Science, Volume 91, 2016, Pages 855-861
- [4] Xiao Zhong, Zeyi Sun, Haoyi Xiong, etc., Learning Curve Analysis Using Intensive Longitudinal and Cluster-Correlated Data, *Procedia Computer Science*, Volume 114, 017, Pages 250-257
- [5] Popchev I., D. Orozova. Towards Big Data Analytics in the E-learning Space *Cybernetics and Information Technologies*, Vol. 19(3), 2019, Pages 16-25.
- [6] Galina Pieva, Decision Analysis for Big Data Platform Selection, *Engineering Sciences*, LVI, 2019, No. 2, pp. 5-17, DOI: 10.7546/EngSci.LVI.19.02.01
- [7] Попчев И., М. Георгиева, Наукометричен анализ в областта на „Големите данни“, Конференция ма БСУ „Хоризонти в развитието на човешките ресурси и знанието“, 2015, Бургас.
- [8] Science Direct – www.sciencedirect.com
- [9] Computer Database (Gale)
- [10] Applied Science & Technology Source (EBSCO)