

# Böyük Verilənlərdən Verilənlər Haqqında Elmə: Fənlərarası Perspektiv

Məkrufə Hacırahimova<sup>1</sup>, Hicran Gözəlova<sup>2</sup>

<sup>1,2</sup>AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

<sup>1</sup>makrufa@science.az, <sup>2</sup>gozalova@yandex.com

**Xülasə** — “Big data” konsepsiyasının populyarlaşması ilə «*data scientist*» daha cəlbedici və perspektivli ixtisasa çevrilmişdir. Məqalədə verilənlər haqqında elmin terminoloji, tarixi aspektləri, mahiyyəti şərh olunmuş, bir fənn kimi bu sahədəki mövcud vəziyyət analiz olunmuşdur. Məqalədə həmçinin verilənlər haqqında elmin fənlərarası perspektivləri araşdırılmışdır.

**Açar sözlər** — böyük verilənlər; verilənlər haqqında elm; verilənlər alimi.

## I. GİRİŞ

Böyük verilənlər (BV) problemi yeni problem deyil. Texniki və texnoloji inkişaf ilə əlaqədar olaraq verilənlərin sürətlə artması və bu verilənlərin emalı problemləri hələ keçən əsrin 40-cı illərinə təsadüf edir [1]. Şübhəsiz o vaxt ilə müqayisədə, son on ildə verilənlərin həcmi kəskin surətdə artmışdır. Bunu IDC, Gartner və s. kimi analitik şirkətlərin hesabatları da təsdiq edir [2].

İnformasiya texnologiyalarının səhiyyə, astronomiya, bioinformatika, nəqliyyat, dövlət idarəçiliyi və s. kimi sahələrdə geniş tətbiqi verilənlər axınının artmasına gətirib çıxarmış, dünyada informasiya bolluğu yaranmışdır. Bunun məntiqi nəticəsi kimi verilənlərin mənimsənilməsi və istifadəsində yeni eranı əks etdirən böyük verilənlər (*ing. big data*) termini meydana çıxmışdır [3].

Verilənlərin həcmnin böyük sürətlə artması həm informasiya istehlakçıları, həm də informasiya emalı texnologiyaları istehsalçıları bu verilənləri real-vaxt ərzində emal edə bilməməsi problemi ilə üz-üzə qoymuşdur. Problem müxtəlif mənbələrdən və müxtəlif formatda (*strukturlaşdırılmış və strukturlaşdırılmamış*) fasiləsiz olaraq generasiya olunan verilənlərin ənənəvi üsul və alətlərin köməyi ilə emal edilməməsindədir. Bəzən “new oil”, bəzən “data mining” və s. kimi dəyərləndirilən BV-nin həqiqi qiyməti qarşılıqlı əlaqəli verilənlər dəstinin analizi nəticəsində biznes tendensiyaları proqnozlaşdırmaq, insan səhhətindəki təhlükələri analiz etmək, terror təhlükələrini müəyyən etmək, axtarışı optimallaşdırılmaq, bioloji və ekoloji tədqiqatlarda və s. dərindən korrelyasiyaları və prinsipləri tapmağa imkan verir. Yəni, yeni nəzəriyyələr, yeni üsullar, analitik alətlər alimlərə və biznes nümayəndələrinə BV-də aşkar şəkildə olmayan biliklərin aşkarlanmasında kömək edə bilər. Bu, verilənlər haqqında elmin əsasını təşkil edir [3,2]. Verilənlər haqqında elm (*ing. data science*) bəzən də datalogiya adlandırılan (*ing. datalogy*)

bu akademik sahə informatikanın bir bölməsi olmaqla rəqəmsal verilənlərin emalı, analizi və təqdim olunması problemlərini öyrənir [4,5,6]. Verilənlər haqqında elm verilənlərdən əlavə biliklərin və qiymətli informasiyanın əldə olunmasıdır. O, riyaziyyat, statistika, obrazların tanınması, biliklər bazası, maşın təlimi və s. kimi fənlərarası (*multidisciplinary*) yanaşmaları təklif edir. Bu, verilənlərin həyat tsiklinin idarə edilməsi, verilənlərin məxfiliyinin təminatı, verilənlərin hesablanması elastikliyi, böyük verilənlərin məkan və zaman xarakterləri və sosial aspektlərinə əsaslanan yeni elmi-tədqiqat sahələrinin yaranmasına səbəb olur. “Big data” konsepsiyasının populyarlaşması ilə o həm elm, həm də biznes strukturları kimi maraqlı tərəflərin diqqət mərkəzinə çevrilmişdir. Belə ki, BV-dən biliklərin aşkarlanmasında yeni keyfiyyətdə kadrlara ehtiyac böyükdür. Bu da “big data” erasında təhsil müəssisələrinin qarşısında duran əsas məsələlərdəndir [3-7]. Verilənlər haqqında elmlə məşğul olan mütəxəssislər verilənlər alimi (*ing. data scientist*) adlanırlar [8]. Onlar bir qayda olaraq, riyaziyyat, statistika və informatikanın əsas elementləri ilə işləmə qabiliyyətinə, həm də müəyyən bir elm sahəsində böyük təcrübəyə malik olurlar.

“Big data” texnologiyaları və verilənlər haqqında elm iqtisadiyyat, biznes, maliyyə, səhiyyə, fizika, astronomiya, geologiya, biologiya, sosial və humanitar elmlər və s. daxil olmaqla müxtəlif sahələrdə tədqiqatların aparılmasına güclü təsir edir. Multidissiplinar baxımdan verilənlər haqqında elm yeni yaranmış sahə kimi bir çox fənləri əhatə edir ki, bu da ona dıgər fənlərlə müqayisədə üstünlük qazandırmaya kömək edir. İşdə də əsas məqsəd akademik fənn kimi “data science” sahəsində mövcud vəziyyəti və perspektivləri araşdırmaqdır.

## II. VERİLƏNLƏR HAQQINDA ELM

“O’Reilly Radar” jurnalında nəşr olunan “Verilənlər haqqında elm nədir?” (What is Data Science?) məqaləsinin müəllifi Mayk Lukidis yazırdı: “Gələcək, verilənləri məhsula çevirə biləcək insan və kompaniyalara məxsus olacaq”. Bu deyim, məşhur bir kəlamı yada salır: “Kim informasiyaya malikdirsə, o da dünyaya sahibdir” [9]. Bu gün bu aforizmə bir az düzəliş etmək olar: “Dünyanı verilənlərə və onların analizi texnologiyalarına malik olanlar idarə edir”. İnformasiyanın alınması üsullarına sahib olmaq verilənlər haqqında elm adlanır. “Data science” termin kimi ilk dəfə 1960-cı illərdə kompüter elmlərinin pionerlərindən sayılan Peter Naur tərəfindən “Datalogy, the science of data and of data processes

and its place in education” adlı məruzəsində səsləndirilmiş [10], fənn kimi isə keçən əsrin 60-cı illərindən formalaşmağa başlamışdır. Bu 1966-cı ildə Elm və Texnika üçün Məlumat Komitəsinin (CODATA – International Council for science: Committee on Data for Science and Technology) təsis edilməsinə təsadüf edir [3,4]. 1974-cü ildə Peter Naur “Concise Survey of Computer Methods in Sweden and the United States” kitabında verilənlərin müasir emal metodlarının icmalını vermiş və verilənlər haqqında elmi rəqəmsal verilənlərin həyat dövrünü (yarandığı andan başqa bilik sahələrinə təqdim olunmaq üçün edilən dəyişikliklərə qədər) öyrənən bir fənn kimi təyin etmişdir. Ancaq, termin 1990-cı illərdə geniş istifadə olunmağa başlamış və 2000-ci illərin əvvəllərindən hamı tərəfindən qəbul edilmişdir. 2001-ci ildə Purdue Universitetinin professoru, statistika, verilənlərin vizuallaşdırılması, maşın təlimi sahəsində tanınmış mütəxəssis Uilyam Klivlend (*William S. Cleveland*) statistik tədqiqatların texniki aspektlərinin inkişaf planı (Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics) adlı tədqiqatında diqqəti “data science” fənninin tədrisi məsələsinə yönəltmişdir [11]. O, planda “data science”-i ayrıca akademik fənn kimi irəli sürmüşdür. Bu da “data scientist” adlanan kadrların hazırlanmasına marağı stimullaşdırmışdır. Təsadüfi deyildir ki, təxminən həmin vaxtdan da CODATA-nın “Data Science Journal”-ı (2002-ci ildən) nəşr olunmağa başlanmışdır.

Burada “data science” ilə bilavasitə əlaqədar bəzi faktları da diqqətə çatdırmaq lazımdır. 1977-ci ildə Statistik Hesablamalar üçün Beynəlxalq Assosiasiyanın (The International Association for Statistical Computing - Section of the ISI) ənənəvi statistik metodlarla müasir kompüter texnologiyaları və ekspertlərin nəticələrini əlaqələndirərək verilənləri informasiyaya və biliyə çevirmək üçün xüsusi seksiya yaradılmışdır [6].

1989-cu ildə Qriqoriy Piatetski-Şapiro verilənlər bazasında biliklərin aşkarlanması (Knowledge Discovery in Databases - KDD) cəmiyyətini yaratmış və 1995-ci ildən hər il keçirilən ACM SIGKDD konfransına çevrilmişdir [6].

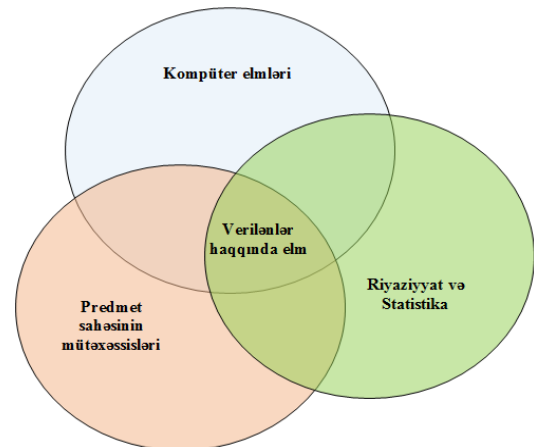
1996-cı ildə verilənlər elmi Təsnifatlandırılmış Cəmiyyətlərin Beynəlxalq Federasiyasının (International Federation of Classification Societies) üzvləri tərəfindən Tokioda, ilk dəfə konfransın adına çıxarılaraq “Data science, classification, and related methods” adlı konfransı keçirilmişdir [6].

1997-ci ildə IBM şirkəti tərəfindən “Biliklərin Aşkarlanması və Verilənlərin İntellektual Analizi” (Knowledge Discovery and Data Mining) jurnalı işıq üzü gördü [12].

### III. FƏNLƏRARASI PERSPEKTİV

“Data science” termini hərfi mənada “verilənlər haqqında elm” kimi tərcümə olunmamalıdır [3]. Çünki, ingilis dilində “science” sözü ancaq “elm” deyil, həm də “bacarıq”, “məharət”, “qabiliyyət” deməkdir. Deməli, bu bilik və bacarıqlara əsaslanan elmdir. Verilənlər haqqında elm ənənəvi

informatikadan başlayaraq riyaziyyata qədər müxtəlif sahələr üzrə bacarıq və vərdiş tələb edir. Bu elmin əsas məqsədi verilənlərdə qanunauyğunluqların tapılması, verilənlərdən ümumiləşdirilmiş halda biliklərin aşkarlanmasıdır. Mayk Lukides verilənlər haqqında elmi verilənləri yalnız informasiyaya deyil, həmçinin məhsula çevirdiyini iddia edir [9]. Bu sahədə vacib olan bacarıqlar adətən, mütəxəssisə lazım olan ümumi fənlərlə informasiya texnologiyalarındakı təcrübələrin və riyazi statistika biliklərinin kəsişməsi – Venn diaqramı [4] vasitəsilə izah olunur (şək.1).



Şəkil 1. Venn diaqramı

Şəkildən də görüldüyü kimi “data science” bir çox elm sahələrinin kəsişməsidir. “Big data” texnologiyalarının tətbiq sahələri artdıqca bu kəsişmələrin sayı da artır. Maşın təlimi, data mining və text mining, süni intellekt, təbii dilin emalı, prediktiv modelləşmə, vizuallaşdırma, təhlükəsizlik və s. kimi elm sahələri daxil olmaqla daha geniş miqyas almaqdadır.

BV-nin potensialından tam istifadə etmək üçün yalnız verilənlər, alətlər və infrastruktura deyil, “nəhəng verilənlər dağını aşmağa” kömək edən çoxlu sayda bacarıq da tələb olunur [13]. Bu sahə ilə məşğul olmaq üçün “data scientist” adlanan yüksək ixtisaslı mütəxəssislərə ehtiyac vardır. Bu mütəxəssislər böyük verilənlərdə qanunauyğunluqların aşkarlanmasında təşkilatlarda aparıcı rola malik olacaqlar.

McKinsey şirkəti 2018-ci ilədək ABŞ-da 440-490 min “data scientist” ixtisaslı mütəxəssisə tələbatın olacağını proqnozlaşdırır. Bu tip proqnozlar bu sahədə təhsil proqramlarının yaranması məsələsini bir qədər də sürətləndirmişdir [14].

ABŞ-ın bir çox ştatlarında və digər inkişaf etmiş dövlətlərin universitetlərində “data science” bir elm sahəsi kimi bakalavr, magistr və doktorluq pilləsində tədris edilir [15]. 2013-cü ildən başlayaraq Dandi Universitetində (Şotlandiya), Oklend Universitetində (Yeni Zelandiya), London İmperial Kollecində, Cənubi Kaliforniya Universitetində, Vaşinqton, Berkli, Nyu-York Universitetlərində və s. verilənlər haqqında proqramlar tədris olunmaqdadır [15]. Belə proqramlar böyük həcmli verilənlərin potensialından maksimum istifadə etmək üçün

hesablama modelləri, modelləşdirmənin və proqnozlaşdırmanın riyazi metodları, arxitektura, müasir proqramlaşdırma metodları, verilənlərin əldə edilməsi, saxlanması və analizi fundamental hazırlığı təmin etməyi nəzərdə tutur. Hazırda bu universitetlərin sayı getdikcə artmaqdadır. Məsələn, Rusiyada Ali İqtisadiyyat Məktəbində, Türkiyədə Sabançı Universitetində və s. magistr pilləsi üzrə proqram tədris olunur. “Data scientist” ixtisası son illərdə daha cəlbədiçə və perspektivli ixtisaslardan biri kimi təqdim edilir.

[15]-də təqdim olunmuş siyahıda adı çəkilən universitetlər xüsusi olaraq böyük verilənlərin analizinə yönəlmiş və bu sahədə magistr pilləsinə yiyələnəməyə nəzərdə tutulmuş fənlər tədris edən ali məktəblərdir. Bu təhsil proqramını öyrənməklə məzunlar böyük verilənləri sərbəst şəkildə idarə edə biləcəklər, ixtisas yönümlü müəssisələr tərəfindən arzuolunan mütəxəssisə çevriləcəklər. Yaxşı mütəxəssis yalnız biznes problemlərinin həlli ilə deyil, müəssisəyə ən çox gəlir gətirən sahəni də düzgün seçməyi bacaracaqdır. Tədris olunan əsas fənlər proqramlaşdırma, alqoritm, verilənlər bazası, riyaziyyat, statistika, verilənlərin analizi, verilənlərin kompüter emalı, paylanmış sistemlər, böyük verilənlər alətləri və verilənlərin vizuallaşdırılmasıdır.

Qeyd edildiyi kimi, “big data” paradigmasının mövcud olması ilə verilənlər haqqında elmə maraq daha da artmışdır. Verilənlər haqqında elm üzrə IEEE International Conference on Data Science and Advanced Analytics, 2011-ci ildən O’Reilly şirkətinin Strata konfransları, EMC-in sammitləri keçirilir.

Bütün bu sadalananlar “data science”-in böyük perspektivlərindən xəbər verir. Böyük verilənləri idarə etmək üçün “verilənlər alimləri”nin əsas bacarığı kod yazmaqdan başlayaraq, analiz etmək və hamının anladığı tərzdə verilənləri vizuallaşdırmaqdır. Müəssisənin müvəffəqiyyəti onlardan birbaşa asılıdır, onların yaratdıqları alətlər isə müəssisə həddlərini aşır. Verilənlər alimləri, məsələn, marketing sahəsində bazarın seqmentlərə bölünməsi, modelləşdirmə, sosial medianın analizi məsələlərini, yaxud maliyyə və sığorta şirkətlərində müxtəlif tətbiqi məsələləri həll edirlər. Şirkətlər, “data scientist” adı altında ayrı-ayrı komandalara və ya şöbələr yaradır ki, onların da tərkibində dörd tip mütəxəssis olur. Bunlar, hər biri biznesdə müxtəlif səviyyəli bacarıqlara – kompüter, statistik və proqramlaşdırma, riyazi biliklərə malik olan “data business person”, “data creative”, “data developer”, “data researcher” qrupundan təşkil olunur. IBM şirkətinin vitse prezidenti Anjul Bhambri “data scientist” haqqında demişdir: “Böyük verilənlər mütəxəssisləri – verilənlərə baxan və tendensiyaları təyin edənlərdir. Onlar, müəssisəyə dəyişiklik

gətirmək üçün öyrənməyi üstün tutan dirçəliş dövrünün şəxsiyyətləridir” [14].

## NƏTİCƏ

Verilənləri biliyin əsas inqridenti hesab etmək olar. Bu baxımdan verilənləri analiz etmək, məlum olmayan bilikləri aşkarlamaq və təqdim etmək vacibdir. Ancaq verilənlərdən biliklərin əldə olunmasında yüksək bilik və bacarığa malik mütəxəssislərə ehtiyac yaranır. Verilənlər artdıqca onları emal etmək üçün informatika və tətbiqi riyaziyyat, statistika və s. kimi elmlərin kəşiməsindən alınan “data science” akademik fənn meydana gəlmişdir. O, bir sıra Amerika və Avropa dövlətlərinin aparıcı universitetlərində müxtəlif səviyyələrdə (*bakalavr, magistr, doktorantura*) tədris olunur. Elmdə fənlərarası istiqamətlərin çoxalması, həmçinin “big data” texnologiyalarının imkanlarından istifadə edən elm sahələrinin sayı artdıqca “data science” daha çox sahənin mütəxəssislərini özündə birləşdirəcəkdir.

## ƏDƏBİYYAT

- [1] Черняк Л. Серьезно о технологиях для Больших Данных, <http://www.osp.ru/os/2014/01/13039646/>
- [2] The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Study report, IDC, December 2012. [www.emc.com/leadership/digital-universe](http://www.emc.com/leadership/digital-universe).
- [3] Əliquliyev R.M., Hacırahimova M. Ş. "Big Data" fenomenoni: problemlər və imkanlar, İnformasiya texnologiyaları problemləri, 2014, №2, səh. 3-16
- [4] [https://ru.wikipedia.org/wiki/Научный\\_Переп](https://ru.wikipedia.org/wiki/Научный_Переп)
- [5] Naur P. «Data science is the science of dealing with data, once they have been established, while the relation of data to what they represent is delegated to other fields and sciences», 1974.
- [6] <http://whatsthebigdata.com/2012/04/26/a-very-short-history-of-data-science/>
- [7] <http://www.datasciencecentral.com/profiles/blogs/17-analytic-disciplines-compared>
- [8] <https://anjulbhambri.wordpress.com/2012/04/27/new-and-exciting-career-opportunities-for-it-professionals-data-scientist/>
- [9] <https://beta.oreilly.com/ideas/what-is-data-science>
- [10] Naur P. Datalogy, the science of data and of data processes and its place in education, *Proc. IFIP Congress*, Edinburgh, Scotland., Amsterdam: North-Holland, 1968, pp. 48-52.
- [11] Cleveland W.S., Statistics Research, Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics, 2001. <http://www.stat.purdue.edu/~wsc/papers/datascience.pdf>
- [12] [http://researcher.ibm.com/researcher/view\\_group\\_pubs.php?grp=144](http://researcher.ibm.com/researcher/view_group_pubs.php?grp=144)
- [13] <http://www.ibmdatahub.com/blog/going-beyond-data-science-toward-analytics-ecosystem-part-1>
- [14] Big data: The next frontier for innovation, competition, and productivity. Analyst report, McKinsey Global Institute, May 2011. <http://www.mckinsey.com/http://datascience.community/colleges>