

INNOVATÍV MÓDSZEREK AZ ADATTUDOMÁNY OKTATÁSÁBAN

INNOVATIVE METHODS IN DATA SCIENCE EDUCATION

Dr. Oparin-Salamon Melinda^{0009-0006-9082-2922 1*}

¹ Menedzsment és Jogi Tanszék, Gazdaságtudományi Kar, Neumann János Egyetem, Magyarország
<https://doi.org/10.47833/2025.2.CSC.002>

Kulcsszavak:

data science
adattudomány
oktatás
statisztika

Keywords:

data science
education
statistics

Cikktörténet:

Beérkezett 2025. augusztus 21.
Átdolgozva 2025. szeptember 5.
Elfogadva 2025. szeptember 9.

Összefoglalás

A digitális korszak fejlődésével az adatok és az adatelemzés szerepe egyre nagyobb jelentőséget kap az üzleti szférában. Az adattudomány – vagy data science – különösen fontos területté vált, mivel létfontosságú lett, hogy a növekvő és egyre összetettebb adathalmazokból értékes következtetéseket lehessen levonni, amelyek támogatják az üzleti döntéshozatalt. Ennek megfelelően a data science készségek oktatása és beépítése a képzési programokba elengedhetetlen. A tanulmány célja az adattudomány oktatási kihívásainak és legfontosabb innovatív módszereinek bemutatása.

Abstract

As the digital era progresses, the role of data and data analysis is gaining increasing importance in the business sector. Data science has become an especially critical field, as it is essential to derive valuable insights from growing and increasingly complex datasets to support business decision-making. Accordingly, teaching data science skills and integrating them into educational programs has become indispensable. This study aims to present the challenges of data science education and highlight its most important innovative methods.

1. Bevezetés

A gyorsan növekvő adatmennyiségnek, valamint a fejlett elemzési eszközök elérhetőségének köszönhetően az adattudomány, vagy más néven data science egyre nagyobb szerepet játszik a vállalatok számára. A data science új lehetőségeket teremt a versenyelőny megszerzésében, mivel lehetővé teszi a piac, az ügyfelek igényeinek, viselkedésének mélyebb feltérképezését, valamint az üzleti folyamatok hatékonyságának növelését. Ennek következtében az adattudományi készségek fejlesztése, oktatási igényként jelent meg az elmúlt években, ami kihívások elé állította az oktatási intézményeket. A tanulmány először az adattudományt definiálja és tömören összeveti a statisztikával, majd az adattudomány oktatásával kapcsolatos igényeket és kihívásokat tárgyalja, valamint bemutatja azokat az innovatív oktatási módszereket, amelyek támogatják a hallgatók felkészítését a szakma gyakorlati kihívásaira.

* Kapcsolattartó szerző.
E-mail cím: salamon.melinda@nje.hu

2. Adattudomány

Gonzales-Argote [1] szerint az adattudomány a számítástechnika és a statisztika egy ága, amely a nagyméretű adathalmazok elemzésére és értelmezésére összpontosít. Gafny és Ben-Zvi [2] megfogalmazásában a data science információs rendszerek vizsgálata valószínűségi számítási módszerekkel, míg ezzel szemben a statisztika szerintük az adatokból tanulás, valamint a bizonytalanság mérése, kontrollálása és kommunikációja.

Tehát a data science egy interdiszciplináris terület, amely a statisztikai módszerek alkalmazását, a számítástechnikai ismereteket és a szakterületi tudást integrálja annak érdekében, hogy értékes betekintéseket nyerjen az adatokból. Bár alapját a statisztika képezi, jelentős eltérések mutatkoznak a klasszikus statisztikai megközelítésekhez képest.

A klasszikus statisztika hagyományosan jól megtervezett mintavétellel gyűjtött adatokat használ a kutatási kérdések megválaszolására. Ezzel szemben a data science gyakran különböző forrásokból származó, sokszor véletlenszerűen vagy más tevékenységek melléktermékeként összegyűjtött adatokat használ fel, ami számos kihívást és eltérést eredményez a statisztikai megközelítésekhez képest. A klasszikus megközelítés elsősorban az értelmezhetőségre és a magyarázóerőre összpontosít, míg az algoritmikus modellezés fő célja a legnagyobb prediktív teljesítményt nyújtó modell kiválasztása, gyakran háttérbe szorítva az interpretálhatóságot. Az adattudomány ezen felül gyakran nem paraméteres modelleket és komplex algoritmusokat használ, amelyek rugalmasabbak a különböző adateloszlások kezelésében, mivel különféle adatstruktúrákat és eloszlásokat képes kezelni anélkül, hogy szigorú feltételezéseket tenne az adatok mögöttes eloszlására. Ezzel szemben a klasszikus statisztika inkább paraméteres modelleket alkalmaz, amelyek meghatározott eloszlási feltételezéseken alapulnak (pl. normál eloszlás). Az eltérő megközelítések különösen szembetűnők a nagy adathalmazok elemzésénél, ahol az adattudomány sokszor hatékonyabb eredményeket ad. [2]

3. Az adattudomány oktatása

Az adattudomány oktatása a fent részletezett jellemzők és statisztikától való eltérések miatt is kihívást jelent az egyetemeken és felsőoktatási intézmények számára. A terület gyors fejlődése miatt a tanterveknek egyszerre kell biztosítaniuk a stabil elméleti alapokat, valamint a gyakorlati, és egyben naprakész készségek fejlesztését. Az oktatás célja nem csupán a technikai ismeretek átadása, hanem komplex kompetenciák kialakítása is, amelyek lehetővé teszik a hallgatók számára a valós problémák megértését, kreatív megközelítését és felelős megoldását.

Az adattudományi képzés több fázisra épül, kezdve az adatgyűjtéstől a történetmesélésig, beleértve az adattisztítást, az adatmodellezést, a vizualizációt és az eredmények bemutatásának lépéseit. Az alapvető programozási nyelvek, mint a Python és az R, illetve az SQL, az adatvizualizációs technikák, valamint a számítástechnikai infrastruktúra ismerete elengedhetetlen az adattudósok számára. Emellett a hallgatóknak fejleszteniük kell problémamegoldó, döntéshozatali, kommunikációs és együttműködési képességeiket is. Az oktatók jelentős szerepet játszanak az új eszközök és stratégiák tanításában, hogy a hallgatók alkalmazkodni tudjanak az adattudomány változó világához, és fontos a valós problémamegoldási forgatókönyvek és vizuális eszközök beépítése is a tananyagba.

Az adattudomány gyorsan fejlődő technológiai és módszertanai folyamatos tantervi felülvizsgálatot igényelnek az oktatási intézmények részéről, hogy az új eszközök és technikák naprakészek maradjanak és lépést tudjanak tartani a vállalati igényekkel. Az adattudomány interdiszciplináris jellege – statisztika, számítástechnika és szakterületi tudás – komplexitást ad a tanterv tervezésének, amelynek fenn kell tartania az összhangot ezen területek között. A hallgatók különböző tudásszintjei miatt személyre szabott megközelítések szükségesek a készség terén meglévő különbségek áthidalásához. Ezen felül kihívást jelent az olyan értékelési módszerek kidolgozása, amelyek pontosan mérik a technikai és elméleti készségeket, mint a kritikus gondolkodást, a kreativitást és a problémamegoldást is. [3], [4]

A hallgatóknak rugalmasnak és képesnek kell lenniük új eszközök és technikák gyors elsajátítására. Fontos, hogy a hallgatók megtanulják, hogy az adattudósoktól elvárt a rugalmasság és az új eszközök, illetve készségek folyamatos elsajátítása. Emellett az etikai kérdések, az adatvédelem és a felelős adatkezelés jelentősége miatt a tantervnek ezeket a témákat is érintenie

kell, hogy a hallgatók felelősségteljesen és etikus módon tudják alkalmazni az adattudományi ismereteket a gyakorlatban. [3], [5], [6]

Az értékelés is meghatározó szerepet játszik az adattudomány oktatásában. Nemcsak a készségek fejlődését támogatja, hanem a hallgatói önbizalmat is erősíti. A self-efficacy felmérések célzott támogatást tesznek lehetővé, a páros összehasonlítások gyorsabb és pontosabb visszajelzést adnak, míg a learning-by-teaching módszer a tudás mélyebb megértését segíti. [7], [8] Emellett az AI-alapú formatív értékelés személyre szabott visszajelzést nyújt, amely bizonyítottan javítja az eredményességet. [9]

4. Innovatív módszerek

A data science oktatásában számos innovatív módszert alkalmaznak, amelyek célja a gyakorlati készségek fejlesztése, a mélyebb megértés elérése, és a tanulók érdeklődésének fenntartása. [10] Ezek a módszerek együtt segíthetnek abban, hogy a hallgatók alkalmazható tudást és készségeket sajátítsanak el a data science terén, és felkészüljenek a gyakorlati kihívásokra, amelyeket a szakma megkíván. Bár Hersh és szerzőtársai (2023) szerint a mögöttes matematikai és programozási ismeretek megértése kevésbé lényeges, mint a módszerek és eredmények megfelelő alkalmazása és értelmezése, ettől függetlenül az adattudomány sokakat elbátortalaníthat a technikai követelmények miatt. A hallgatók elköteleződésének és motivációjának fenntartása elengedhetetlen a megtartás érdekében. Innovatív tanítási módszerek, például projektalapú tanulás, csoportos feladatok, és valós alkalmazások segíthetnek az érdeklődés fenntartásában. Továbbá, a vizuális eszközök, mint a grafikus adatvizualizációs programok, szintén hozzájárulhatnak a tanulás élményéhez. Megfelelő támogató szolgáltatások, például mentorálás és tanácsadás biztosítása kulcsfontosságú a hallgatók számára, hogy leküzdjék a nehézségeket és sikeresek legyenek az adattudományi képzésekben.

A Wright és szerzőtársai [11] bemutatják a Virginiai Egyetem Adattudományi Iskolája által kidolgozott Adattudományi Alapszak tantervét, amelynek a célja, hogy a hallgatókat átfogó ismeretekkel ruházzák fel annak érdekében, hogy sikeresen dolgozhassanak adattudományi csapatokban. Kiemelik az átfogó oktatási megközelítés fontosságát és a hatékony, STEM-orientált innovatív oktatási módszerek alkalmazását. Ezek közé tartozik, hogy olyan értékelési rendszert dolgoztak ki, amely a készségek elsajátítására összpontosít, szemben a hagyományos osztályozási módszerekkel. Interaktív előadások, valamint a vendégelőadások tartása is az új módszertan részét képezi. Míg az előbbi a hallgatók bevonásáról szól, addig a másik esetén iparági szakértők osztják meg tapasztalataikat és engednek betekintést a munkájukba. Szintén új elem a gamifikációs (játékosító) laborok alkalmazása, amely során játékos elemeket építenek a gyakorlati feladatokba és ezáltal növelik a tanulás hatékonyságát, illetve az elköteleződést.

Dusen és szerzőtársai [10] a UC Berkeley Foundation of Data Science programját mutatják be, amely a számítástechnika társadalmi hatásainak integrálásával, a hallgatók érdeklődésére építve fejleszti a számítástechnikai készségeket. A program céljai között szerepel a sokszínűség növelése a számítástechnikát tanuló hallgatók körében, erős etikai alapok biztosítása, valamint kritikai gondolkodás ösztönzése a statisztikai technikák alkalmazásában. Yan és Davis [12] szintén átfogó áttekintést nyújtanak egy alapszintű adattudományi kurzus felépítéséről, tanulmányukban kiemelik a gyakorlati alkalmazások fontosságát, valamint az iparági szempontból releváns készségek integrálását, és data science életciklus alapvető fogalmát.

Az adattudomány oktatásában hatékony módszernek tekinthetők az olyan aktív tanulási módszerek, amelyek egyaránt fejlesztik a technikai készségeket és a kommunikációs kompetenciát. A projektalapú tanulás bizonyítottan erősíti a hallgatók problémamegoldó képességét, valamint elősegíti a mélyebb fogalmi megértést. [10] Emellett az interdiszciplináris elemek – például etika, tudományos kommunikáció és számítástechnikai alapismeretek – integrálása javítja a hallgatók felkészültségét a valós adatproblémák kezelésére. A mesterséges intelligencia oktatásba való bevonásával személyre szabott, azonnali visszajelzés nyújtására van lehetőség, ami pedig a gyorsabb készségfejlesztést támogatja. Az innovatív módszerek közül az alábbiak azok, amelyeket több szerző is jelentős elemként azonosít [7], [8], [9], [13], [14], [15]:

1. Projektek és gyakorlati tapasztalatok: A projektalapú tanulás során a hallgatók saját adattudományi projektjeiken dolgozhatnak, ami lehetőséget ad az adatgyűjtés, az

adattisztítás, az adatmodellezés, és vizualizáció folyamatainak gyakorlati megtapasztalására. Ez növeli az önálló problémamegoldási képességeket és fejleszti a csapatmunkát, illetve összhangban van az adattudomány felfedező jellegével. Azonban az egyes projektek túl szűk területre fókuszálhatnak, ezért a hallgatók nem kapnak teljes módszertani áttekintést.

2. Hackathonok és versenyek: Az adat alapú hackathonok, illetve versenyek lehetőséget adnak a hallgatóknak, hogy időre dolgozzanak egy valós vagy valóság-hű problémán, és közben valódi kihívásokkal szembesüljenek. Ezek az események növelik a résztvevők kreativitását és gyors gondolkodását, valamint elősegítik a csapatok közti együttműködést. Korlátja, hogy az idő rövidsége miatt kevés lehetőség nyílik az alapos adatelőkészítésre és az eredmények értelmezésére. Így a látványos eredmények kerülnek előtérbe, a módszertani mélységgel szemben.
3. Vizualizációs és interaktív eszközök: Az interaktív adatelemző és vizualizációs platformok, (mint például a Tableau, Power BI, vagy a Jupyter Notebook) segítségével a hallgatók valós időben láthatják az elemzési folyamat eredményeit, mivel lépésről lépésre végeznek számítási műveleteket a programkódok fokozatos írásával, így könnyebben megértik az adatok közti összefüggéseket. Ez egyben a módszer hátránya is lehet, mivel a túlságosan felhasználóbarát eszközök a mögöttes statisztikai és algoritmikus folyamatok megértését hátráltathatják.
4. Szimuláció és játékosítás (gamifikáció) : Szimulációs modellek és játékosítási technikák alkalmazása (például virtuális adatgyűjtési környezetekben vagy probléma alapú feladatokban) jobban bevonja a hallgatókat és lehetőséget nyújt számukra, hogy játékos formában sajátítsák el a szükséges készségeket. A szimulációk azonban sokszor túlságosan leegyszerűsítik a valós környezet összetettségét.
5. Esettanulmányok és valós adatfeladatok: Az esettanulmány alapú megközelítéssel a hallgatók olyan valós problémákat tanulmányoznak, amelyekkel az iparban dolgozó szakemberek is találkoznak. Valós adatfeladatokkal dolgozva a tanulók megtanulják, hogyan kezeljék a valódi adatproblémákat, és kialakítanak egyfajta kritikus gondolkodást. A módszer korlátja lehet, hogy az ipari adatok sokszor nem nyilvánosak vagy anonimizáltak, ami csökkenti az életszerűséget vagy olyan adatmennyiségek feldolgozására van szükség, ami nehezen illeszthető szűk kurzuskeretekbe.
6. Önálló tanulás és online források használata: Az online tananyagok, például videók, videókurzusok lehetőséget adnak arra, hogy a résztvevők saját tempójukban tanuljanak. A lépésről lépésre kidolgozott, programkóddal kiegészített (pl. Python) útmutatók alkalmazása segíti a hallgatókat az adattisztítás, az adatfeldolgozás és a vizualizáció hatékony elsajátításában. Ez a módszer magas önfegyelmet és motivációt igényel a hallgatók részéről.
7. Automatizált feladatellenőrzés és visszajelzés: Az automatizált feladatértékelő rendszerek lehetővé teszik a hallgatók számára, hogy azonnali visszajelzést kapjanak, és az oktatók számára is könnyebbé teszik a tanulási folyamat nyomon követését. Az automatizált értékelés viszont nem mindig méri a gondolkodási folyamatot és kreatív megoldásokat; a hallgatók a teszt lefutására optimalizálhatnak a mélyebb megértés helyett.

Bár ezek az innovációk ígéretesek, továbbra is kihívást jelent, hogy minden hallgató egyenlő hozzáférést kapjon az erőforrásokhoz és támogatáshoz, valamint az, hogy az oktatók is rendelkezzenek a szükséges készségekkel, erőforrásokkal. Az adattudomány oktatása számos kihívással szembesül, amelyek hatással vannak a képzési programok minőségére és hatékonyságára. A tantervi hiányosságok (például a bevezető kurzusok hiánya és az előre definiált átjárási útvonalak kidolgozatlansága) akadályozzák a hallgatók előrehaladását.

5. Összegzés

Az adattudomány folyamatos fejlődésével az oktatási intézményeknek innovatív oktatási módszerekkel kell felkészíteniük a hallgatókat a data science képzéseiket annak érdekében, hogy a leendő adattudósok képesek legyenek az üzleti kihívásoknak megfelelni. Ilyenek a projekt alapú tanulás, amely önálló problémamegoldási képességeket fejleszt, és a hackathonok, amelyek valós

problémákon dolgoztatják a hallgatókat, ösztönözve a kreativitást és csapatmunkát. Az interaktív adatelemző eszközök, a szimulációk és a játékosítás segítik az adatok közti összefüggések megértését, míg az esettanulmányok és valós adatfeladatok a kritikai gondolkodást fejlesztik. Az online tananyagok és az automatizált visszajelzési rendszerek pedig lehetőséget adnak a hallgatóknak, hogy saját tempójukban tanuljanak és azonnali visszajelzést kapjanak. Az olyan környezet kialakítása, amely ösztönzi a fejlődést és az együttműködést, lehetővé teszi, hogy a hallgatók alkalmazásorientált módon sajátítsák el az adattudomány alapjait és technikáit. Az aktív tanulási módszerek, például a problémamegoldó feladatok, a valós adatkészletek feldolgozása és a projektalapú tanulás egyaránt támogatják a tananyag rugalmas és releváns átadását. Az egyetemeken és az oktatók közös erőfeszítése szükséges ahhoz, hogy olyan programokat hozzanak létre, amelyek reagálnak a hallgatók és az iparágak állandóan változó igényeire, biztosítva, hogy az oktatási környezet a lehető legjobban készítse fel a hallgatókat a gyakorlati kihívásokra és a folyamatos fejlődésre, tanulásra.

Irodalomjegyzék

- [1] D. Gonzalez Argote, „Thematic Specialization of Institutions with Academic Programs in the Field of Data Science”, *Data and Metadata*, köt. 2, o. 24, márc. 2023, doi: 10.56294/dm202324.
- [2] R. Gafny és Ben-Zvi, Dani, „Reimagining data education: Bridging between classical statistics and data science”, in *Reasoning with data models and modeling in the big data era*, 2022.
- [3] M. Oparin-Salamon, „A data science oktatási kihívásai: a statisztikai tudás és az adattudományi készségek integrálása”, in *Statisztika+ Nightingale konferencia előadásai*, ME-GEMI., Miskolc, 2024. [Online]. Elérhető: <https://gemi.uni-miskolc.hu/files/19055/WP04.pdf#page=41> ISBN 978-963-358-351-7
- [4] M. Birkenkrahe, „Teaching Data Science with Literate Programming Tools”, *Digital*, köt. 3, sz. 3, o. 232–250, szept. 2023, doi: 10.3390/digital3030015.
- [5] T. Donoghue, B. Voytek, és S. E. Ellis, „Teaching Creative and Practical Data Science at Scale”, *Journal of Statistics and Data Science Education*, köt. 29, sz. sup1, o. S27–S39, jan. 2021, doi: 10.1080/10691898.2020.1860725.
- [6] B. Memarian és T. Doleck, „Data science pedagogical tools and practices: A systematic literature review”, *Educ Inf Technol*, köt. 29, sz. 7, o. 8179–8201, máj. 2024, doi: 10.1007/s10639-023-12102-y.
- [7] S. Malallah, E. Osiobe, és Shamir, „Data Science Self-Efficacy Assessment Tools: A Foundational Guide to Evaluating Progress”, 2024. július 1., *Open Science Framework*. doi: 10.31219/osf.io/xqy5s.
- [8] L. K. Chen és J. Thai, „Assessment-via-Teaching: Exploring an Alternative Assessment Strategy in Undergraduate Introductory Data Science Course”, in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, Portland OR USA: ACM, márc. 2024, o. 1598–1599. doi: 10.1145/3626253.3635516.
- [9] I. Letteri és P. Vittorini, „Enhancing Student Feedback in Data Science Education: Harnessing the Power of AI-Generated Approaches”, 2024. december 4., *In Review*. doi: 10.21203/rs.3.rs-5505390/v1.
- [10] E. Van Dusen, J. DeNero, és K. Usovich, „Innovation in Data Science Education”, in *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*, Providence RI USA: ACM, márc. 2022, o. 1202–1202. doi: 10.1145/3478432.3499154.
- [11] B. Wright, P. Alonzi, és A. Riveria, „The Future of Data Science Education”, 2024, *arXiv*. doi: 10.48550/ARXIV.2407.11824.
- [12] D. Yan és G. E. Davis, „A First Course in Data Science”, *Journal of Statistics Education*, köt. 27, sz. 2, o. 99–109, máj. 2019, doi: 10.1080/10691898.2019.1623136.
- [13] F. Martinez-Plumed és J. Hernández-Orallo, „Training Data Scientists Through Project-Based Learning”, *IEEE R. Iberoamericana Tecnologías Aprendizaje*, köt. 18, sz. 3, o. 295–304, aug. 2023, doi: 10.1109/RITA.2023.3302954.
- [14] K. Oyetade, T. Zuva, és A. Harmse, „Evaluation of the impact of hackathons in education”, *Cogent Education*, köt. 11, sz. 1, o. 2392420, dec. 2024, doi: 10.1080/2331186X.2024.2392420.
- [15] M. Kalogiannakis, S. Papadakis, és A.-I. Zourmpakis, „Gamification in Science Education. A Systematic Review of the Literature”, *Education Sciences*, köt. 11, sz. 1, o. 22, jan. 2021, doi: 10.3390/educsci11010022.