

KATTA MALUMOTLAR DATA SCIENCE**Tojimamatov Israil Nurmatovich***Farg'ona Davlat Universiteti amaliy matematika
va informatika kafedrası o'qituvchisi***Topvoldiyev Asadbek Nushonboy o'g'li***Farg'ona davlat universiteti 3-kurs talabasi*topvoldievasadbek@gmail.com

Annotatsiya: *Ushbu maqola katta ma'lumotlar (Big Data) va ma'lumotlar ilmi (Data Science) sohasining dolzarb masalalarini tahlil qilishga bag'ishlangan. Tadqiqotda katta ma'lumotlarning asosiy xususiyatlari, qo'llanilish sohalari, innovatsion ma'lumotlarni tahlil qilish usullari, texnologiyalarning rivojlanishi va ilmiy tadqiqotlar uchun imkoniyatlar ko'rib chiqiladi. Tadqiqotning maqsadi katta ma'lumotlar va Data Science sohalarining amaliy ahamiyatini, ularda yuzaga kelgan yangi trendlarni aniqlash, shuningdek, kelajakdagi ilmiy tadqiqotlar va texnologiyalar uchun istiqbollarni belgilashdir.*

Kalit so'zlar: *Katta ma'lumotlar, Data Science, mashinani o'rganish, sun'iy intellekt, tahlil, algoritmlar, tuzilmagan ma'lumotlar, real vaqt tahlili, kiberxavfsizlik, blokcheyn, prediktiv tahlil, tabiiy tilni qayta ishlash (NLP), kompyuter ko'rish, cloud computing, distributed systems, NoSQL.*

Аннотация: *Эта статья посвящена анализу актуальных вопросов в области больших данных (Big Data) и науки о данных (Data Science). В исследовании рассматриваются основные характеристики больших данных, области их применения, инновационные методы анализа данных, развитие технологий и возможности для научных исследований. Цель исследования заключается в определении практического значения больших данных и науки о данных, выявлении новых тенденций в этих областях, а также в установлении перспектив для будущих научных исследований и технологий.*

Ключевые слова: *Большие данные, Наука о данных, Машинное обучение, Искусственный интеллект, Анализ, Алгоритмы, Неструктурированные данные, Анализ в реальном времени, Кибербезопасность, Блокчейн, Прогностический анализ, Обработка естественного языка (NLP), Компьютерное зрение, Облачные вычисления, Распределённые системы, NoSQL.*

Abstract: *This article is dedicated to analyzing the current issues in the field of Big Data and Data Science. The research discusses the main characteristics of Big Data, areas of application, innovative methods of data analysis, the development of technologies, and opportunities for scientific research. The aim of the research is to determine the practical significance of Big Data and Data Science, identify new*

trends in these fields, and outline the prospects for future scientific research and technologies.

Keywords: *Big Data, Data Science, Machine Learning, Artificial Intelligence, Analysis, Algorithms, Unstructured Data, Real-Time Analysis, Cybersecurity, Blockchain, Predictive Analytics, Natural Language Processing (NLP), Computer Vision, Cloud Computing, Distributed Systems, NoSQL.*

Zamonaviy texnologiyalar va raqamli inqilob ma'lumotlarning shakli, hajmi va tezligi jihatidan yangi davrga olib keldi. Ushbu ma'lumotlar turli sohalarda, jumladan, iqtisodiyot, sog'liqni saqlash, ekologiya, ta'lim va hatto ijtimoiy munosabatlarda qo'llaniladi. Katta ma'lumotlar (Big Data) konsepsiyasi ma'lumotlarni saqlash, qayta ishlash va tahlil qilishda yuzaga kelayotgan yangi muammolar va imkoniyatlarni anglatadi. Ushbu yangi ma'lumotlar ulkan hajmdagi strukturalanmagan (unstructured) va yarim strukturalangan ma'lumotlardan iborat bo'lib, ularni samarali tahlil qilishda Data Science (ma'lumotlar ilmi) asosiy rol o'ynaydi. Data Science sohasida sun'iy intellekt (AI), mashinani o'rganish (ML), chuqur o'rganish (Deep Learning), tabiiy tilni qayta ishlash (NLP), va statistik tahlil metodlari qo'llaniladi. Katta ma'lumotlar bilan ishlashning asosiy maqsadi nafaqat ma'lumotni samarali saqlash, balki undan foydali bilimlarni olishdir. Ushbu maqolada katta ma'lumotlar, Data Science va ularning ilmiy tadqiqotlar uchun qadrli imkoniyatlari batafsil tahlil qilinadi.

Amaliy misollar

Katta ma'lumotlar amaliyoti ko'plab sohalarda o'z aksini topadi. Misol uchun, Amazon kompaniyasi katta ma'lumotlardan foydalanib, mijozlarning xarid qilish odatlarini tahlil qiladi. Bu kompaniya foydalanuvchilarning qiziqishlariga mos ravishda tavsiyalar berish orqali savdoni oshiradi. Amazonning «Sizga yoqishi mumkin» bo'limi, katta ma'lumotlardan foydalanishning ajoyib misolidir. Netflix ham katta ma'lumotlardan foydalanadigan yana bir misoldir. U foydalanuvchilarning tomosha qilish odatlarini tahlil qilib, yangi filmlar va seriallarni yaratish yoki mavjudlarini taklif qilishda yordam beradi. Netflix foydalanuvchilarining qanday kontentni ko'rishini bilib, ularni qoniqtirish uchun yangi kontent yaratishga imkon beradi. Sog'liqni saqlash sohasida katta ma'lumotlar tahlili bemorlarning holatini yaxshilashda qo'llaniladi. Masalan, sog'liqni saqlash tashkilotlari katta ma'lumotlardan foydalanib, kasalliklarni oldini olish va davolash strategiyalarini ishlab chiqadilar. Bemorlarning tarixiy ma'lumotlarini tahlil qilib, ularning salomatlik holatini prognoz qilish va shaxsiylashtirilgan davolash rejalarini ishlab chiqish mumkin. Sport sohasida ham katta ma'lumotlar muhim rol o'ynaydi. Sport jamoalari o'yinchilarining statistikalarini tahlil qilib, ularning o'yinlarini yaxshilashga harakat qiladilar. Misol uchun, NBA jamoalari o'yinchilarining harakatlarini kuzatish uchun sensorli texnologiyalarni ishlatadilar va bu ma'lumotlarni tahlil qilib, o'yin strategiyalarini ishlab chiqadilar.

Kamchiliklar

Katta ma'lumotlar bilan bog'liq muammolar ko'p. Bularning eng muhimlari orasida maxfiylik va xavfsizlik masalalari mavjud. Katta ma'lumotlar to'planganda, shaxsiy ma'lumotlar xavf ostiga tushishi mumkin. Masalan, ijtimoiy tarmoqlarda foydalanuvchilarning shaxsiy hayoti haqida to'plangan ma'lumotlar noto'g'ri qo'llanilishi mumkin. Shuning uchun kompaniyalar va tashkilotlar foydalanuvchilarning maxfiyligini himoya qilish uchun qoidalar va standartlarga rioya qilishlari zarur. Bundan tashqari, katta ma'lumotlarni tahlil qilishda kutilmagan natijalar yoki xatolar paydo bo'lishi mumkin. Ma'lumotlarni noto'g'ri talqin qilish yoki noto'g'ri model yaratish natijasida noaniq yoki noto'g'ri qarorlar qabul qilinishi mumkin. Bu esa biznes jarayonlariga salbiy ta'sir ko'rsatishi mumkin. Ma'lumotlarni boshqarish ham katta muammo hisoblanadi. Katta hajmdagi ma'lumotlarni saqlash, qayta ishlash va tahlil qilish uchun kuchli infratuzilma talab etiladi. Ko'p hollarda kompaniyalar yangi texnologiyalarni joriy etish yoki mavjud tizimlarini yangilashda muammolarga duch keladilar.

Yangiliklar va tendensiyalar

Katta ma'lumotlar sohasidagi yangiliklar va tendensiyalar doimiy ravishda o'zgarib bormoqda. Sun'iy intellekt (AI) va mashinani o'rganish (ML) kabi texnologiyalar katta ma'lumotlarni tahlil qilishni yanada samarali qiladi. Ushbu texnologiyalar yordamida kompaniyalar katta hajmdagi ma'lumotlarni tezda qayta ishlash va ulardan foydali natijalar olish imkoniyatiga ega bo'ladi.

Bulutli hisoblash texnologiyalari ham katta ma'lumotlarni boshqarishda muhim rol o'ynaydi. Bulutli xizmatlar orqali kompaniyalar o'z ma'lumotlarini saqlash va qayta ishlash uchun kerakli resurslarga ega bo'lishadi. Bu esa ularga xarajatlarni kamaytirish va operatsion samaradorlikni oshirish imkonini beradi.

Shuningdek, real vaqt rejimida katta ma'lumotlarni tahlil qilish tendensiyasi ham kuchaymoqda. Kompaniyalar real vaqt rejimida foydalanuvchilar bilan muloqotda bo'lish va ularga mos xizmatlarni taklif qilish uchun katta ma'lumotlardan foydalanmoqdalar. Masalan, onlayn do'konlarda xarid qilish jarayonida mijozlarga real vaqt rejimida tavsiyalar berish mumkin.

Kelajak istiqbollari

Katta ma'lumotlar sohasidagi kelajak istiqbollari juda umidvor. Sun'iy intellekt va mashinani o'rganish texnologiyalarining rivojlanishi bilan katta ma'lumotlardan foydalanish yanada kengayadi. Kompaniyalar yanada murakkab tahlillarni amalga oshirish va yanada aniqroq prognozlar qilish imkoniyatiga ega bo'lishadi.

Shuningdek, katta ma'lumotlarning shaxsiylashtirilgan xizmatlarga aylanishi kutilmoqda. Foydalanuvchilar uchun individual ehtiyojlariga moslashtirilgan xizmatlar yaratish uchun katta ma'lumotlardan foydalaniladi. Bu esa mijozlarga yanada yaxshi tajriba taqdim etadi.

Bundan tashqari, katta ma'lumotlar bilan bog'liq qonunchilik va etik masalalar ham kelajakda muhim ahamiyat kasb etadi. Maxfiylikni himoya qilish va shaxsiy ma'lumotlarni boshqarish bo'yicha yangi qonunlar qabul qilinishi kutilmoqda.

Adabiyotlar Tahlili

Ilmiy adabiyotlarda katta ma'lumotlar va Data Science sohasidagi muhim yo'nalishlar quyidagicha ajratib ko'rsatiladi:

Ma'lumotlarni boshqarish tizimlari arxitekturasini va optimallashtirish masalalari Katta hajmdagi ma'lumotlarni samarali saqlash, qayta ishlash va boshqarish uchun ma'lumotlar bazalari va tizimlari arxitekturasini talab etiladi. Masalan, Hadoop, Spark, NoSQL bazalari kabi texnologiyalar katta hajmdagi ma'lumotlarni tez va samarali qayta ishlashga yordam beradi. Shu bilan birga, data lake yoki data warehouse arxitekturasini orqali ma'lumotlarni integratsiya qilish va tahlil qilish samaradorligini oshirish mumkin.

Ma'lumotlar xavfsizligi va maxfiylikni ta'minlash texnologiyalari Katta ma'lumotlar xavfsizligi muammosi doimiy ravishda dolzarb bo'lib qoladi. Ma'lumotlarning xavfsizligini ta'minlash uchun kriptografiya, ma'lumotlarni shifrlash, blokcheyn texnologiyasi va kiberxavfsizlik metodlari qo'llaniladi. Shu bilan birga, GDPR (General Data Protection Regulation) kabi huquqiy va axborot xavfsizligi masalalari katta ma'lumotlarni boshqarishda muhim rol o'ynaydi.

Katta hajmdagi tuzilmagan ma'lumotlarni tahlil qilish algoritmlari Katta ma'lumotlarning ko'pchiligi tuzilmagan (unstructured) shaklda bo'ladi, masalan, matnlar, rasmlar, video va audio fayllar. Bunday ma'lumotlarni tahlil qilish uchun tabiiy tilni qayta ishlash (NLP) va kompyuter ko'rish (Computer Vision) texnologiyalari samarali qo'llaniladi. Bu texnologiyalar yordamida ijtimoiy tarmoqlardan, forumlardan va bloglardan olingan ma'lumotlardan foydali bilimlar olish mumkin.

Sun'iy intellekt va mashinani o'rganish usullarining katta ma'lumotlar tahliliga integratsiyasi

Sun'iy intellekt va mashinani o'rganish texnologiyalari katta ma'lumotlarni tahlil qilishda muhim rol o'ynaydi. Deep learning usullari yordamida katta hajmdagi ma'lumotlar orqali yuqori samaradorlikka erishish mumkin. Shu bilan birga, supervised learning, unsupervised learning, reinforcement learning kabi mashinani o'rganish metodlari katta ma'lumotlar tahlilining samaradorligini oshiradi.

Natijalar va muhokama

Tadqiqot davomida katta ma'lumotlarning asosiy xususiyatlari va ularni tahlil qilishda yuzaga keladigan muammolarni aniqlash mumkin bo'ldi:

Hajm

Katta ma'lumotlarning hajmi doimiy ravishda o'sib bormoqda, bu esa saqlash va qayta ishlashda yangi texnologiyalarni joriy etishni talab qiladi. Cloud computing va

distributed systems kabi texnologiyalar katta hajmdagi ma'lumotlarni saqlash va tahlil qilishda samarali vositalar hisoblanadi.

Tezlik

Ma'lumotlar real vaqt rejimida yig'ilmoqda, bu esa ularni tezkor tahlil qilishni talab qiladi. Stream processing texnologiyalarini (masalan, Apache Kafka, Flink) qo'llash orqali ma'lumotlarni real vaqtda tahlil qilish mumkin. Buning natijasida tezkor qarorlar qabul qilish imkoniyatlari oshadi.

Xilma-xillik

Katta ma'lumotlar turli shakllarda bo'ladi, masalan, matn, rasm, video, sensorli ma'lumotlar va boshqalar. Ularni tahlil qilishda NoSQL bazalari, graph databases va data lakes texnologiyalari yordamida samarali yondashuvlar qo'llaniladi.

Haqiqat

Katta ma'lumotlarning ishonchliligi va to'g'riligi ta'minlanishi zarur. Data cleansing (ma'lumotlarni tozalash) va data validation (tekshirish) usullari yordamida ma'lumotlarni sifatli va ishonchli qilish mumkin. Bunda AI-based anomaly detection texnologiyalaridan foydalanish samarali bo'ladi.

Qiymat

Katta ma'lumotlardan olingan bilimlarni biznes va ilmiy tadqiqotlarda foydali qilib qo'llash mumkin. Predictive analytics, prescriptive analytics va descriptive analytics kabi metodlar orqali ma'lumotlardan yuqori qiymat olish mumkin.

Xulosa va tavsiyalar

Katta ma'lumotlar va Data Science sohalari nafaqat iqtisodiyotda, balki ilm-fan, sog'liqni saqlash, ekologiya, ta'lim va boshqa sohalarda ham katta ahamiyatga ega. Katta ma'lumotlarni samarali tahlil qilish uchun ilg'or texnologiyalar va metodlar zarur. Data Science sohasining rivojlanishi bilan sun'iy tatistic, mashinani o'rganish, chuqur o'rganish va tatistic tahlil texnologiyalari katta ma'lumotlarni tahlil qilishda yuqori samaradorlikka erishishga yordam beradi.

FOYDALANILGAN ADABIYOTLAR:

1.Davenport, T. H., & Dyché, J. (2013). Big Data in Big Companies. International Institute for Analytics. Ushbu kitob katta ma'lumotlar sohasidagi qo'llanilishi va ularni ishlatishda yuzaga keladigan muammolarni tahlil qiladi.

2.Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt. Katta ma'lumotlarning jamiyat va biznesga ta'siri, shuningdek, uning ilmiy tadqiqotlar va texnologiyalar uchun imkoniyatlari muhokama qilinadi.

3.Bengio, Y., Courville, A., & Vincent, P. (2013). Learning Deep Architectures for AI. Foundations and Trends® in Machine Learning, 2(1), 1-127.

Chuqur o'rganish va sun'iy intellekt sohasidagi asosiy usullar va algoritmlar tahlil qilinadi.

4.Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. Statistik tahlil va mashinani o'rganish asosida katta ma'lumotlar bilan ishlash bo'yicha ilmiy ishlar.

5.Koller, D., & Friedman, N. (2009). Probabilistic Graphical Models: Principles and Techniques. MIT Press. Ma'lumotlar ilmi va mashinani o'rganish asosida graf tarmoqlari va ehtimollarni tahlil qilish usullari.

6.Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM, 51(1), 107-113. Katta ma'lumotlar tahlili va saqlanishi uchun samarali MapReduce texnologiyasining yondashuvlari.

7.Zikopoulos, P., Eaton, C., DeRoest, M., & Deutsch, T. (2012). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill. Hadoop va stream processing texnologiyalari haqida chuqur tahlil.

8.Khan, S., & Hussain, M. (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing, and Predicting with Big Data. Wiley. Data Science va katta ma'lumotlarni tahlil qilish bo'yicha yangiliklar va amaliy qo'llanmalar.

9.Gartner, Inc. (2013). The 2013 Gartner Magic Quadrant for Business Intelligence and Analytics Platforms. Gartner Research. Katta ma'lumotlar tahlili va biznes intellekti bo'yicha so'nggi tendensiyalar.

10.Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. Mobile Networks and Applications, 19(2), 171-209.