

MA'LUMOTLARNI QAYTA ISHLASH VA TAHLIL QILISH

Umarov Bekzod Azizovich

Farg'ona davlat universiteti, Amaliy matematika va informatika kafedrası, o'qituvchi Gmail: ubaumarov@mail.ru

Ibrohimjonov Ma'rufjon Hokimjon o'g'li

Farg'ona davlat universiteti, talabasi
Gmail: marufjonibrohimjonov0@gmail.com

Annotatsiya: Ma'lumot olimlari katta ma'lumotlarni to'plab, tahlil qilib, sharhlashadi va ular orqali naqshlarni aniqlash, tushunchaga ega bo'lish, bashorat qilish hamda harakat rejalarini ishlab chiqadilar. Katta ma'lumotlar tahlili ma'lumotlarning xilma-xilligi, hajmi va olish tezligi kabi omillarga bog'liq. Ma'lumot olimlari strukturaviy va strukturaviy bo'lmagan ma'lumotlar bilan ishlaydilar. Strukturaviy ma'lumotlar odatda raqamlar va so'zlardan tashkil topgan bo'lsa, strukturaviy bo'lmagan ma'lumotlar esa matnli hujjatlar, ijtimoiy media va video ma'lumotlaridan iborat bo'lishi mumkin. Ma'lumotlar olimlari tahlil qilish va modellarni yaratish uchun dasturlash, statistika, ehtimollar nazariyasi, mashina o'rganish kabi ko'nikmalarga ega bo'lishlari kerak. Ular o'zlarining tahlillarini va ma'lumotlarni biznes natijalariga aylantirishda, o'zlarining domen bilimlarini ham qo'llaydilar.

Kalit so'zlar: Ma'lumot olimi, katta ma'lumotlar, strukturaviy ma'lumotlar, strukturaviy bo'lmagan ma'lumotlar, dasturlash, statistika, ehtimollar nazariyasi, mashina o'rganish, vizualizatsiya, bashorat qilish, domen bilimlari, real vaqtli tahlil.

Annotation: Data scientists collect, analyze, and interpret large datasets to identify patterns, gain insights, make predictions, and develop action plans. Big data analytics is defined by the variety, volume, and velocity of data that surpass traditional data processing capabilities. Data scientists work with both structured data (such as spreadsheets) and unstructured data (such as text documents, social media, and video data). Their skills include programming, mathematics, statistics, probability theory, machine learning, and domain knowledge. These skills enable them to solve industry-specific problems, optimize business outcomes, and transform raw data into actionable insights.

Keywords: Data scientist, big data, structured data, unstructured data, programming, statistics, probability theory, machine learning, visualization, prediction, domain knowledge, real-time analytics.

Аннотация: Учёные данных собирают, анализируют и интерпретируют большие объёмы данных для выявления закономерностей, получения инсайтов, предсказаний и разработки планов действий. Анализ больших данных определяется такими характеристиками, как разнообразие, объём и скорость

данных, которые превосходят возможности традиционных методов обработки данных. Учёные данных работают как с структурированными данными (например, электронные таблицы), так и с неструктурированными данными (такими как текстовые документы, социальные медиа и видеоданные). Они обладают навыками программирования, математики, статистики, теории вероятностей, машинного обучения и отраслевых знаний. Эти навыки помогают им решать специфические задачи отрасли, оптимизировать бизнес-результаты и превращать сырые данные в полезную информацию.

Ключевые слова: Учёный данных, большие данные, структурированные данные, неструктурированные данные, программирование, статистика, теория вероятностей, машинное обучение, визуализация, прогнозирование, отраслевые знания, реальное время анализ.

Ma'lumotlarni qayta ishlash va tahlil qilishning ta'rifi

Ma'lumotlar fani - bu bilimlarni yaratishga qaratilgan ma'lumotlarni ilmiy o'rganish. Ushbu soha ko'plab fanlarni birlashtiradi va ularning usullarini katta ma'lumotlar to'plamlaridan bilim olish uchun qo'llaydi, shunda ular asosli qarorlar qabul qilish va bashoratlarni yaratish uchun ishlatilishi mumkin. Ma'lumotni qazib olish nafaqat ma'lumotlar bo'yicha olimlarni, balki ma'lumotlar tahlilchilari, ma'lumotlar arxitektorlari, ma'lumotlar muhandislari, statistiklar, ma'lumotlar bazasi ma'murlari va biznes tahlilchilarni ham o'z ichiga oladi.

Ma'lumotlar hajmi eksponent tarzda o'sib bormoqda va ma'lumotlarni qayta ishlash va tahlil qilishga bo'lgan ehtiyoj jadal o'sib bormoqda va kompaniyalar daromadni oshirish va innovatsiyalarni rivojlantirish uchun tobora ko'proq tahlilga bog'liq. Misol uchun, biznes aloqalari tobora raqamli bo'lib, ko'proq ma'lumotlar yaratiladi. Shunday qilib, tajribalarni shaxsiylashtirish, mijozlar ehtiyojini qondirish, xizmat ko'rsatishni yaxshilash, yangi va takomillashtirilgan mahsulotlarni ishlab chiqish va sotishni oshirish bo'yicha tushunchaga ega bo'lish uchun yangi imkoniyatlar mavjud. Bundan tashqari, biznes olamida va undan tashqarida ma'lumotlar fani ba'zan eng murakkab global muammolarni hal qilishga yordam beradi.



Ma'lumot olimi nima qiladi?

Ma'lumot olimi naqshlarni aniqlash, tushunchaga ega bo'lish, bashorat qilish va harakat rejalarini ishlab chiqish uchun katta ma'lumotlarni to'playdi, tahlil qiladi va sharhlaydi. Katta ma'lumotlar tahlili Katta ma'lumotlar xilma-xilligi, hajmi va olish tezligi oldingi ma'lumotlarni boshqarish usullarida mavjud bo'lgan qayta ishlash imkoniyatlaridan yuqori bo'lgan ma'lumotlar to'plami sifatida belgilanishi mumkin. Ma'lumotlar olimlari katta ma'lumotlarning ko'p turlaridan foydalanadilar. Ulardan ba'zilari quyida keltirilgan.

Strukturaviy ma'lumotlar odatda qatorlar va ustunlarga ajratiladi va so'zlar va raqamlarni (ismlar, sanalar, kredit karta ma'lumotlari va boshqalar) o'z ichiga oladi. Masalan, kommunal sohada ishlaydigan ma'lumot olimi xarajatlarni kamaytirish imkoniyatlarini izlash va uskunaning ishdan chiqishiga olib kelishi mumkin bo'lgan naqshlarni aniqlash uchun elektr energiyasi ishlab chiqarish va foydalanish ma'lumotlarining elektron jadvallarini tahlil qilishi mumkin.

Tarkibi bo'lmagan ma'lumotlar hech qanday tarzda tartibga solinmagan va matnli hujjat fayllari, ijtimoiy tarmoqlar va mobil qurilmalar ma'lumotlari, veb-sayt tarkibi va video ma'lumotlarini o'z ichiga olishi mumkin. Misol uchun, chakana ma'lumotlar olimi mijozlar tajribasini qanday yaxshilash bo'yicha savollarga javob berish uchun tuzilmagan qo'ng'iroq markazi eslatmalarini, elektron pochta xabarlarini, so'rovnomalarni va ijtimoiy media ma'lumotlarini tahlil qilishi mumkin.

Bundan tashqari, ma'lumotlar to'plamining xarakteristikalarini miqdoriy (tuzilmali raqamli ma'lumotlar) yoki sifatli yoki toifali ma'lumotlar (sonli qiymatlar bilan ifodalanmagan va toifalarga guruhlanishi mumkin bo'lgan ma'lumotlar) sifatida tavsiflanishi mumkin. Mutaxassislar uchun qaysi turdagi ma'lumotlar bilan ishlayotganligini bilish juda muhim, chunki bu ishlatiladigan tahlil turlarini va ma'lumotlarni vizualizatsiya qilish uchun mos bo'lgan grafik turlarini aniqlaydi.

Ushbu barcha turdagi ma'lumotlardan bilim olish uchun ma'lumotlar olimlari quyida sanab o'tilgan sohalarida o'z ko'nikmalaridan foydalanadilar.

Dasturlash Ma'lumotlar bo'yicha olimlar korporativ ma'lumotlar bazasidan ma'lumot olish uchun Julia, R yoki Python kabi tillarda so'rovlar yozadilar. Ko'pchilik Python-dan foydalanishni afzal ko'radi, chunki u ma'lumotlarni qayta ishlash va tahlil qilish uchun tayyor modullarni taklif qiladi va hatto kodlash tajribasi bo'lmagan odamlar tomonidan ham oson o'rganilishi va ishlatilishi mumkin.

Matematika, statistika va ehtimollar nazariyasi. Ushbu ko'nikmalar ma'lumotlarni tahlil qilish, gipotezalarni sinab ko'rish va ma'lum turdagi naqshlarni tan olishga o'rgatilgan fayllar bo'lgan mashinani o'rganish modellarini yaratishga yordam beradi. Ma'lumotlar olimlari ma'lumotlardagi munosabatlarni aniqlash, ma'lumotlar haqida bashorat qilish va muammolarni hal qilish uchun o'qitilgan mashinani o'rganish modellaridan foydalanadilar. Modellarini noldan yaratish va o'qitish o'rniga ular avtomatlashtirilgan mashinani o'rganishdan foydalanishlari va

tayyor ishlaydigan mashinani o'rganish modellariga kirishlari mumkin.

Mavzu bo'yicha bilim . Ma'lumotlarni biznes natijalarini yaxshilashga yordam beradigan amaliy tahlilga aylantirish uchun ma'lumot olimi domen bilimiga ega bo'lishi kerak, ya'ni sanoatda va o'z kompaniyasida nima sodir bo'layotganini tushunishi kerak. Quyida ma'lumotlar olimlari o'zlarining domen bilimlarini sanoat muammolarini hal qilish uchun qanday qo'llashlari mumkinligiga misollar keltirilgan.



Ma'lumotlarni qayta ishlash va tahlil qilish jarayonlari

O'z loyihalarini amalga oshirish uchun ma'lumotlar olimlari quyidagi jarayonga o'xshash jarayonga rioya qilishlari kerak.

1. Biznes muammosini aniqlang

Ma'lumotlar bo'yicha olimlar hal qilinishi kerak bo'lgan muammo yoki javob berilishi kerak bo'lgan savolni, shuningdek, loyiha maqsadlari va yechim talablarini aniq belgilash uchun manfaatdor tomonlar bilan ishlaydi.

2. Analitik yondashuvni aniqlash

Biznes muammosini tushunish asosida ma'lumotlar olimi analitik yondashuvni tanlaydi:

tavsiflovchi, hozirgi holat haqida ko'proq ma'lumot olish;
diagnostika, nima sodir bo'layotganini va nima uchun ekanligini tushunish;
bashorat qilish, kelajakda nima bo'lishini bashorat qilish;
muammoni qanday hal qilishni tushunish uchun ko'rsatma.

3. Ma'lumotlarni qabul qilish

Ma'lumot olimi kerakli natijaga erishish uchun zarur bo'lgan ma'lumotlarni aniqlaydi va oladi. Bu ma'lumotlar bazalarini so'rash, veb-saytlardan ma'lumotlarni yig'ish (vab qirqish) yoki fayllardan ma'lumot olish bo'lishi mumkin. Ma'lumotlar ichki foydalanish uchun mavjud bo'lishi mumkin, lekin sotib olish kerak bo'lishi mumkin. Ba'zi hollarda tashkilotlar loyihani muvaffaqiyatli boshlash uchun yangi ma'lumotlarni to'plashlari kerak bo'lishi mumkin.

4. Ma'lumotlarni tozalash

Odatda, bu bosqich eng ko'p vaqtni oladi. Modellashtirish uchun ma'lumotlar to'plamini yaratish uchun ma'lumot olimi barcha ma'lumotlarni bir formatga aylantiradi, ularni tartibga soladi, kerak bo'lmagan narsalarni olib tashlaydi va etishmayotganini almashtiradi.

5. Ma'lumotlarni o'rganish

Ma'lumotlarni tozalagandan so'ng, ma'lumot olimi uni o'rganishni boshlaydi. U ma'lumotlar elementlari qanday bog'liqligini va ular va taxmin qilingan qiymatlar (yorliqlar deb ataladi) o'rtasida qanday statistik munosabatlar kuzatilganligini tushunish uchun statistik tahlil usullaridan foydalanadi. Prognoz yorlig'i miqdoriy qiymat bo'lishi mumkin, masalan, kelajakdagi xarajatlar yoki parvozning bir necha daqiqada kechikishi davomiyligi.

O'rganish va tayyorlash odatda interaktiv tahlil va ma'lumotlarni vizualizatsiya qilishni o'z ichiga oladi. Odatda, Python va R kabi tillar ushbu maqsadlar uchun, interaktiv vositalar va ushbu vazifalar uchun maxsus mo'ljallangan muhitlarda qo'llaniladi. Ma'lumotlarni ko'rish uchun ishlatiladigan skriptlar odatda Jupyter Notebook kabi maxsus muhitlarda joylashtiriladi. Ushbu vositalar ma'lumotlar olimlariga ma'lumotlarni va hujjatlarni dasturiy ravishda ko'rish va ular kashf etgan naqshlarni almashish imkonini beradi.

6. Ma'lumotlarni modellashtirish

Ma'lumot olimi retsept yoki tavsifiy modelni yaratadi va o'qitadi, so'ngra savolga javob berishi yoki biznes muammosini hal qilish uchun uni sinovdan o'tkazadi va baholaydi. Oddiy ma'noda model - bu kirishni qabul qiladigan va chiqishni ishlab chiqaradigan kod qismidir. Mashinani o'rganish modelini yaratish algoritmi tanlash, uni ma'lumotlar bilan ta'minlash va giperparametrlarni sozlashni o'z ichiga oladi.

Giperparametrlar - bu modelni o'qitish jarayonini boshqarish uchun ishlatiladigan sozlanishi parametrlar. Masalan, neyron tarmoqlarda ma'lumot olimi yashirin qatlamlar sonini va har bir qatlamdagi tugunlar sonini aniqlaydi. Giperparametrlarni sozlash (yoki optimallashtirish) eng yaxshi ishlashni ta'minlaydigan giperparametr konfiguratsiyasini topish jarayonidir.

Siz eshitadigan keng tarqalgan savol: "Qaysi mashinani o'rganish algoritmidan foydalanishim kerak?" Mashinani o'rganish algoritmi ma'lumotlar to'plamini modelga aylantiradi. Algoritmi tanlash, birinchi navbatda, ma'lumotlarni qayta ishlash va tahlil qilish stsenariysining ikki jihatiga bog'liq:

O'tgan ma'lumotlar bo'yicha modelni o'rgatish orqali ma'lumot olimi qanday biznes savoliga javob bermoqchi?

Ma'lumotlarni qayta ishlash va tahlil qilish stsenariysi qanday talablarni qo'yadi: aniqlik, o'qitish vaqti, chiziqlilik, parametrlar soni, funktsiyalar soni va boshqalar?

Ushbu savollarga javob berish uchun Azure Machine Learning ko'p toifali qarorlar o'rmoni, tavsiya tizimlari, neyron tarmog'ining regressiyasi, ko'p sinfli

neyron tarmog'i va k-vositalari klasteri kabi keng qamrovli algoritmlar portfelini taqdim etadi. Har bir algoritm mashinani o'rganishning ma'lum bir turini hal qilish uchun mo'ljallangan. Bundan tashqari, Azure Machine Learning algoritmlari varaqasi biznes savolingizga javob berish uchun to'g'ri algoritmni tanlashga yordam beradi.

7. Modelni joylashtirish

Ma'lumot olimi yakuniy modelni hujjatlar bilan taqdim etadi va sinovdan o'tgandan so'ng, tashkilot undan faol foydalanishi uchun yangi ma'lumotlar to'plamini ishlab chiqarishga joylashtiradi. Joylashtirilgan modeldan olingan bashoratlardan biznes qarorlarini qabul qilishda foydalanish mumkin.

8. Vizualizatsiya va natijalar taqdimoti

Microsoft Power BI, Tableau, Apache Superset va Metabase kabi vizualizatsiya vositalari ma'lumotlaringizni o'rganishni va texnik bo'lmagan odamlarga natijalaringizni tushunishga yordam beradigan chiroyli vizualizatsiya yaratishni osonlashtiradi.

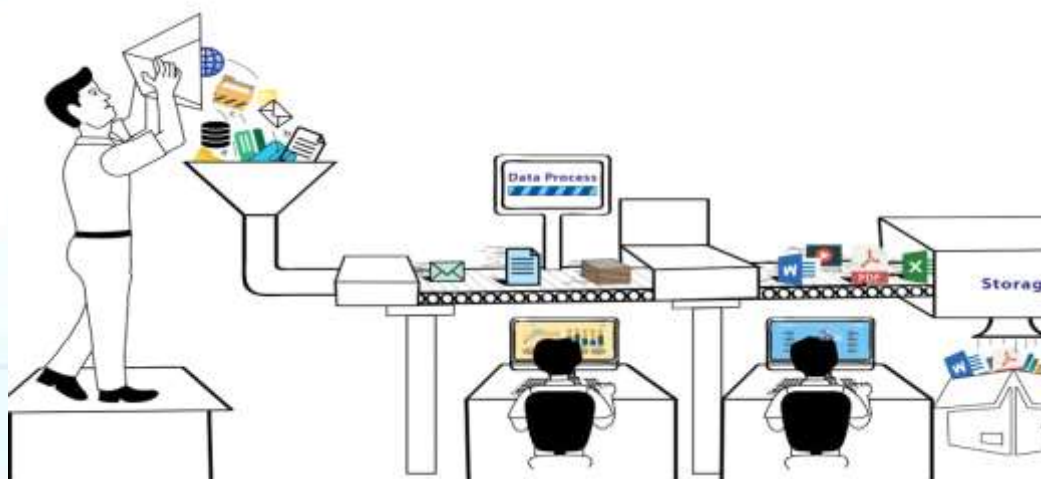
Bundan tashqari, ma'lumotlar olimlari Zeppelin noutbuklari kabi onlayn daftarlardan ma'lumotlarni to'plash, kashf qilish, tahlil qilish, vizualizatsiya qilish va hamkorlik qilishda yordam berishlari mumkin.

Ma'lumotlarni qayta ishlash va tahlil qilish usullari

Ma'lumotlar bo'yicha olimlar statistik jihatdan haqiqiy tahlillarni ishlab chiqarish uchun gipoteza sinovi, klasterlash, omil tahlili va regressiya tahlili kabi statistik usullardan foydalanadilar.

Ma'lumotlar fanining hujjatlari

Ma'lumotlar fanining hujjatlari loyiha va sohaga qarab farq qilsa-da, u odatda ma'lumotlar qayerdan kelgani va qanday o'zgartirilganligi haqidagi ma'lumotlarni o'z ichiga oladi. Bu boshqa jamoa a'zolariga ishlayotganda ushbu ma'lumotlardan samarali foydalanishga yordam beradi. Masalan, hujjatlar biznes tahlilchilariga vizualizatsiya vositalaridan foydalangan holda ma'lumotlar to'plamini sharhlashda yordam beradi.



Ma'lumotlarni qayta ishlash va tahlil qilish hujjatlari quyidagi turlarga bo'linadi.

Loyiha rejalari - loyihaaning biznes maqsadlarini, baholash ko'rsatkichlarini, resurslarini, vaqt jadvallarini va byudjetini aniqlash imkonini beradi.

Ma'lumotlar fanidan foydalanuvchi hikoyalari ma'lumotlar fanlari bo'yicha loyihalar uchun g'oyalarni ishlab chiqishga yordam beradi. Mutaxassis manfaatdor tomonlarning nuqtai nazarini aks ettiruvchi hikoyani yozadi, ular nimaga erishmoqchi ekanliklarini va nima uchun ushbu loyihani talab qilishlarini ko'rsatadilar.

Ma'lumotlarni qayta ishlash va tahlil qilish modeli hujjatlari - ma'lumotlar to'plamini, eksperimental dizaynni va algoritmlarni hujjatlashtirishga imkon beradi.

Qo'llab-quvvatlovchi tizim hujjatlari - foydalanuvchi qo'llanmalari, tizimga texnik xizmat ko'rsatish infratuzilmasi hujjatlari va kod hujjatlarini o'z ichiga oladi.

Xulosa: Ma'lumot olimi katta ma'lumotlarni tahlil qilib, ma'lum naqshlar, aloqalar va tendentsiyalarni aniqlash orqali biznes qarorlarini qo'llab-quvvatlaydi. Ular strukturaviy va strukturaviy bo'lmagan ma'lumotlarni ishlatadilar, masalan, elektron jadvallar, ijtimoiy tarmoqlar yoki video materiallar. Ma'lumot olimlari tahlil qilishda dasturlash, statistika va mashina o'rganish kabi ko'nikmalarni qo'llaydilar. Ushbu bilimlarni sanoatdagi maxsus muammolarni hal qilishda va biznes jarayonlarini takomillashtirishda samarali ishlatishadi. Ma'lumotlarni vizualizatsiya qilish va bashorat qilish orqali muammolarni tez va aniq aniqlash imkonini beradi.

FOYDALANILGAN ADABIYOTLAR:

1. "Internet of Things: A Comprehensive Approach to Enabling Technologies and Applications" - Rajkumar Buyya, Amir Vahid Dastjerdi (2016)

2. "Architecting the Cloud: Design Decisions for Cloud Computing Service Models" - Michael J. Kavis (2017)

3. "Cloud Computing and the Internet of Things: A Comprehensive Survey" - N. Kumar, R. S. Sahu, A. M. Al-Hawari, et al. (2019)

4. "The Internet of Things: Opportunities and Challenges for Health Care" - T.J. O'Rourke, D.J. Houghton (2020)

5. "Building the Future: Big Teaming for the 21st Century" - Mark W. Johnson (2018)

6. "Cloud Computing for Data-Intensive Applications: Concepts, Technology and Applications" - C.Y. Chan, H.P.H.Sharma (2021)

7. "Internet of Things in Smart Technologies for Sustainable Urban Development" - N.A.B.Abdul Rahman, Y.M.A.Rahman (2021)

8. "Cloud-Based Internet of Things Frameworks for Smart Home Automation Systems: A Survey and Comparative Analysis" - M.A.Gumel, Y.P.Kumar (2022)