

Дәріс 1. Деректер туралы ғылымына кіріспе.

Кіріспе.

Деректер туралы ғылым(Data Science) дегеніміз не?

Деректердің рөлі.

Деректерді зерттеушілер деген кімдер?

Мысалдар мен бағыттар.

Кіріспе.

Бүгінгі әлем деректер үстемдігі дәуірінде өмір сүреді. Осы заманғы қоғамдағы бірде-бір күн немесе бірде-бір секунд әртүрлі құрылғылардан деректер жинап, сақтап, өңдемей өтпейді.

Интернет беттері әр қолданушының шертулері, іздеулері және сатып алулары арқылы ақпаратты үшінші тараптардың ондағаны пайдаланып, жарнаманы бағыттауға ниеттенген. Смартфондар ауа райы немесе браузер жарнамалары сияқты қосымшалардан қалып жататын деректер арқылы біздің орналасқан жерімізді және интернет сатып алулар қызығушылықтарымызды анықтау мақсатында қолданылады. Сонымен қатар, бұл ақылды терминалдар, сондай-ақ әлеуметтік желілер ақпарат алмасуға арналған онлайн-платформалар осының бәрі біріктірілген ақпараттық экожүйеге өңделіп, ол күшті байланыстылыққа ие болып, адамның өмір сүру өлшеміне сай деректер құрастырады.

Интернет адамзат білімінің орасан зор кеңістігі болып табылатын ең үлкен деректер торабына айналды, ол фильмдер мен музыка, ғылыми жаңалықтар, спорт нәтижелері, саясат және қоғамдық трендтерді бірге қаржы және маркетинг ақпаратын да қамтиды.

Бұл деректердің шексіз теңізінде біздің мақсатымыз қандай? Адамзаттың міндеті осы ақпараттың барлығын мәніне қарай саралап, оны шынайы білім мен шешімдерге айналдыру. Бұл қажеттілік деректер ғылымының дамуымен байланысты. Деректер ғылымы — бұл математика, статистика, бағдарламалау және салалар бойынша білімнің тоғысқан жері. Ол деректерден пайдалы үлгілерді, болжауларды және ақпараттандырылған шешімдерді көрсетуді көмекке алады. Бұл ғылымның мақсаты — деректерді ақпаратқа айналдыру, бұл жинақталған сандар мен мәтіндер емес, мағыналы әрі пайдалы заттар.

Бүгінгі таңда деректер стратегиялық актив ретінде қарастырылады. Ұйымдар деректерге нарық трендтерін талдау және сұранысты болжау үшін арқа сүйейді. Дәрігерлер науқастардың деректерін қолдана отырып, ауруларды алдын ала емдеу және алдын алу әдістерін әзірлеуге тырысады. Зерттеушілер климаттың өзгеруін, экология

мәселелерін және күрделі табиғи құбылыстарды модельдеуде үлкен деректерді қолданады. Деректер адамдардың өмір сүру сапасын жақсартуға, жол кептелістерін болдырмауға және қоғамды қауіпсіздендіруге ықпал етеді.

Осы дәрісте біз деректер ғылымының қандай екенін және қоғамдағы оның пішінін жақынырақ қарастырамыз. Деректер ғылымы саласының негізгі салалары және құралдары туралы білу арқылы, ақпараттың қалай жиналатыны, өңделетіні және нәтижесінде қалай білім мен интуицияға айналатыны туралы сөйлесеміз. Бұл қазіргі заманғы мамандар үшін өте маңызды тақырып, өйткені деректерді өңдеу қабілеті ХХІ ғасырдағы олардың ең маңызды кәсіби дағдыларының бірі болып табылады.

Деректер туралы ғылым(Data Science) дегеніміз не?

Деректер ғылымы — бұл деректерден білім мен инсайтты алу үшін математикалық, статистикалық және компьютерлік әдістерді қолданатын пәнаралық сала. Қарапайым сөзбен айтқанда, деректер ғылымы — бұл қолданбалы күш-жігер саласындағы деректерді жинауда, өңдеуде, талдауда және соңында алынған деректер жиынтығын түсініп, ақпараттандырылған шешімдер қабылдауда қолданылады. Бұл ғылым компьютерлік ғылым, статистика, жасанды интеллект және бизнес-талдау негізінде дамыды. Оның негізгі мақсаты — үлкен көлемдегі құрылымдық және құрылымдық емес деректерден үлгілер мен қарым-қатынастарды тану. Бірақ менің ойымша "Деректер ғылымы" тек деректермен жұмыс жасауды ғана емес, сондай-ақ болжауды, әрекеттер жөнінде кеңестер беруді және ұсыныс жасауды да білдіреді.

Жалпы, деректер ғылымы жұмыс процесінің өз кезеңдері бар:

1. **Деректерді жинау** — әртүрлі көздерден деректер жинау: интернет беттері, сенсорлар және әлеуметтік желілер немесе сауалнамалар.
2. **Деректерді тазалау және өңдеу** — қажетсіз немесе қателерді ақпараттан тазалау, форматтарды нормалау.
3. **Талдау және модельдеу** — статистикалық және машиналық оқыту алгоритмдерін қолдану арқылы қорытынды алу.
4. **Нәтижелерді интерпретациялау және көрсету** — нәтижелерді басқаларға түсінікті түрде жасау, теңдеулер немесе диаграммаларды қолдану.
5. **Шешімдер қабылдау** — талдауды әрекетке көшіру кезеңі.

Мысалы, электрондық коммерция фирмалары сатып алу тарихын талдап, келесі қандай өнім ұсынылатынын болжау мүмкіндігін бағалайды. Мысалы, дәрігерлер медициналық деректерді қолдана отырып, денсаулықтың нашарлауының

ықтималдылығын анықтайды, ал мемлекеттік агенттіктер деректер ғылымы құралдарын қаладағы жол қозғалысын жақсарту үшін қолданады.

Қысқаша айтқанда, деректер ғылымы — бұл деректер арқылы шынайылыққа жаңа тәсіл. Сіз оған қарап, өткенді түсініп алатын сияқты, қазіргі уақытта сипаттау және болашаққа қарау.

Деректердің рөлі.

Біздің қоғамымыздың ең қымбат активі — деректер. Егер қоғамның дамуының алғашқы ғасырларында экономикалық және индустриялық әдеттер табиғи ресурстардан және шикізаттан алынған болса, қазір ақпаратты жеңіп алу және оны ең ыңғайлы түрде пайдалануға мүмкіндігі бар екені ұтады. Деректер жаңа дәуірдің "мұнайы" болып табылады, өйткені оны дұрыс өңдей алатын ұйымның үлкен артықшылығы бар. Сондықтан деректер технологиялық фетиш ретінде қолданылады және стратегиялық және қоғамдық маңызға ие.

Деректер қазіргі өмірдің әрбір бұрышына және қуыс-қуыс түстеніп отыр. Бизнеске олар бізге нарықтағы сұранысты болжауға, бәсекелестерімнің не істеп жатқанын және тұтынушылардың қалай әрекет ететіні туралы түсінуге көмектеседі. Біз студенттерді жетістіктерін бағалау, оқу процесін дербестендіру және сапалы білім алу үшін деректерді пайдалануды үйретеміз. Денсаулық сақтау саласында деректер аурудың ерте сатыларын анықтауға, эпидемиялардың қалай әрі қаншалықты тез таралатынын болжауға және емдеу әдістерінің тиімділігін түсінуге көмектеседі. Ғылымда деректер күрделі табиғи немесе әлеуметтік құбылыстардағы үлгілер мен құрылымдарды ашуға көмектеседі.

Деректер қоғамдық басқарудың бас көзі. Оларға саясат құру және қоғамның экономикалық жағдайын бағалау үшін арқа сүйенеміз. Деректер шешім қабылдау сапасын арттыруда маңызды рөл атқарады. Деректерге негізделген шешім ішкі сезімдерге емес, нақты дәлелдерге және талдауға негізделген. Ол ұйымның сенімін және жұмысын арттырып, қауіптерге ұшырамау ықтималдығын азайтады. Сонымен қатар, жаңалықтың басты көзі болып табылады. AI, машиналық оқыту, үлкен деректер, автоматтандыру — олардың барлығына деректер қажет. Деректердің талдаулары әлеуметтік өмірді жеңілдетіп жіберді, жаңа өнімдер мен қызметтерді жасауды ынталандырды. Навигациялық жүйелер арқасында тамаша жол қозғалысы, онлайн дүкендерден тұтынушыларға бағытталған ұсыныстар мен көптеген басқа мысалдар әртүрлі ақпарат түрлерінің талдауының нәтижесі.

Бірақ неғұрлым деректер көп болса, соғұрлым проблемалар да көп болады: жеке кеңістіктің бұзылуынан қорғаныс, деректер сапасының қателіктері және этика сияқты мәселелер маңызды болды. Сондықтан кез-келген деректердің өңделуі жауапкершілік

және мөлдірлік қағидаларына негізделуі тиіс. Деректерді қорғау және сапалы дайындалуы біздің уақыттың ең қажетті талаптары ретінде анықталады.

Деректерді зерттеушілер деген кімдер?

Деректерді зерттеушілер (Data scientists) - үлкен көлемдегі ақпарат арқылы инсайттарды анықтап, пішіндеу және болжау бойынша сарапшылар болып табылады. Олар проблемаларды шешу үшін модельдер жасайды, әсіресе математикалық, статистикалық және бағдарламалық білімдерді қолдану арқылы. Қысқаша, деректер ғалымы ақпарат жасайтын адам болады. Деректер талдау маманының ықпал ету аймағы бірнеше негізгі саланы қамтиды:

- **Деректер тазалау және трансформациялау** — әртүрлі дереккөздерден деректер алу және қолайлы форматтарда дайындау.
- **Деректерді қазып алу және модельдеу** — статистика және машиналық оқыту арқылы деректерде жасырушы байланыстарды анықтау.
- **Болжамдық талдау және шешімдерді қолдау** — болашақ өзгерістерді болжау және байқауы бойынша ұсыныстар жасау.
- **Нәтижелерді визуализациялау** — бұл ақпаратты графиктер, диаграммалар және интерактивті элементтерімен кестелерде оңай сіңімді түрде көрсету.

Деректер ғалымдары тек техникалық білімдері бар мамандар емес. Олар логикалық ойлау, аналитикалық ойлау және сыни түсіну қабілеттеріне ие болуы керек. Негізгі мақсат — деректерден кейбір мағына алу, содан кейін бұл инсайтты құндылыққа айналдыру — бизнес немесе ғылым немесе қоғам үшін.

Деректер ғалымының рөлі өте кең. Мысалға бизнес саласында, мұндай алгоритмдер тұтынушылардың жүріс-тұрысын талдайды және сату және маркетинг стратегияларын оңтайластыру; денсаулық сақтау саласында, олар аурулардың таралуын болжайды және ем беру шешімдерін жекешелендіру; қаржы жүйелерінде алаяқтарды анықтау немесе қауіптерді азайту үшін модельдер жасалған; ал білім беру және ғылымда осы модельдер күрделі үрдістерді имитациялауда көмектеседі және жаңа ғылыми теорияларды дамытады.

Күн сайын, деректер ғалымдары Python, R және SQL тәрізді тілдерге сүйеніп байттарға ақпараттар арасынан өтіп отырып, TensorFlow немесе Scikit-learn сияқты машиналық үйрену кітапханаларымен және визуализация бағдарламалық құралдарымен (Power BI, Tableau немесе Matplotlib) бірге жұмыс істей отырып, қолданады. Сонымен қатар, олар үлкен деректермен жұмыс істеу үшін жиі AWS бұлттық қызметтерін пайдаланады.

Мысалдар мен бағыттар.

Деректер ғылымы қазір адам өмірінің барлық салаларында қолданылады. Ол тек техникалық немесе ғылыми деңгейде шектелмейді, бизнес, медицина, білім беру, қаржы, көлік және тіпті мемлекет басқару жүйесінің аясында да іске асырылады. Төменде деректер ғылымының басты практикалары және одан әрі егжей-тегжейлі мысалдары келтірілген:

1. **Бизнес және маркетинг** Қазіргі әлемде бизнес деректерге тең келеді. Компаниялар деректер ғылымын өздерінің тұтынушылары, нарық трендтерін болжау және табысты арттыру үшін қолданады. Мысалға ретінде Amazon (сатып алулар) немесе Netflix (фильмдер мен сериалдар) немесе Spotify (музыка) сияқты үлкен компаниялар деректер ғылымын қолданушылардың әрекеттеріне негізделген ұсыныс жүйелерін құрастыру үшін пайдаланады. Бұл жүйелер бұрынғы таңдаулар, іздеу тарихы және әрекеттерді талдап, келесі қандай нәрсе ұнауы мүмкін екенін болжайды. Деректер ғылымы компанияларға өз жарнамаларын неғұрлым нақты мақсатқа бағыттауда, аз шығынмен көмек көрсетеді. Маркетологтар тұтынушыларын бірнеше топқа бөліп, әр топқа арнайы стратегия дайындау арқылы шығара алады. Сондықтан бұл бизнес операцияларды және тұтынушылардың қанағаттануын арттырады.

2. **Денсаулық сақтау** Деректер ғылымы арқылы ең көп бұзылған салалардың бірі — денсаулық сақтау саласы. Денсаулық сақтау саласында деректер науқастың денсаулығын бақылау арқылы аурулардың ерте диагнозын анықтау және емдеу сапасын жақсартады. Мысалы, жасанды интеллект алгоритмдері рентген және МРІ суреттерін қарап, ерте онкологиялық немесе жүрек ауруларына байланысты алғашқы сигналдарды автоматты түрде анықтай алады. Сонымен қатар, науқастардың медициналық жазбалары және генетикалық ақпараттары жекеленген емдеу жоспарын әзірлеуде қолданады — бұл дербестелген медицина.

3. **Білім беру** Білім беруде деректер ғылымы студенттердің жетістіктерін қадағалау және білім беру жүйесінің сапасын жаңғыртуға негізделген оқу процесін ұсынады. Мысалы, онлайн платформалар (Coursera, edX, Khan Academy) студенттердің оқу әрекеттерін қадағалап, олар қиындық келтіретін тақырыптарды анықтап, қосымша ресурстар ұсынады. Бұл "белсенді оқу" деп аталатын оқыту әдісі. Қатысу, тапсырмаларды орындау және тест нәтижелері университеттер мен колледждердегі білім беру бағдарламаларының тиімділігімен байланыстырып көрсетеді. Сондықтан деректер ғылымы білім сапасын, әділдігін және өзектілігін жақсартуға жұмыс етеді.

4. **Қаржы және банк секторы** Қаржы секторында банктер деректер ғылымын қауіптерді азайтуға және тұтынушыларға сенімді қызмет көрсетуге көмектеседі. Несие берушілер клиенттің қаржы тарихы мен транзакцияларын талдап, несиелік сенімділікті бағалай алады. Машиналық оқыту алгоритмдері арқылы сіз автоматты түрде күдікті транзакцияларды анықтап және жариялай аласыз. Мысалы, егер картаның иесі әдеттен тыс елде немесе уақытта төлем жасаса, жүйе автоматты түрде оларға ескерту береді. Деректер компанияларға нарық трендтерін болжауға және сәтті активтерді басқару модельдерін құруға көмектеседі. Қаржы тұрақтылығы мен қауіпсіздігі. Бұл бөлімдерде деректер ғылымы маңызды.

5. **Өнеркәсіп және өндіріс** Өндіріс 4.0 дәуірінде өндіріс процесі тек деректер арқылы басқарылады. Мыңдаған және одан да көп жаңа зауыттар мен фабрикалардағы датчиктердің ішінде, автомобиль жасау немесе энергия өндіру қондырғыларындағы машиналардың күйі, температура, қысым немесе өнім сапасы тұрақты тексеріледі. Инженерлер деректерді қарап отырып, ақауларды ерте ұстап және өндіріс тоқтағанға дейін жөндеу жұмыстарын жүргізе алады. Бұл болжамдық техникалық қызмет деп аталады. Өйткені деректерді өндіру процесін дәлдеген және ресурстарды үнемдеген. Мысалы, энергияны пайдалану аномалиялар үшін тексеріледі. Өнім сапасы жоғарылайды, ал шығындар азаяды.

6. **Көлік және логистика** Көлік және логистика саласында деректер ғылымы оңтайлы жолдарды құру және уақыт пен шығынды үнемдеуде қолданылады. Мысалы, Uber, Bolt, Yandex Go сияқты платформалар жолаушылар мен жүргізушілерді мүмкін болатын қысқа сәйкес келетін нақты уақыт деректеріне сәйкес байланыстырады. Логистика компаниялары (DHL немесе Amazon Logistics) жеткізу маршруттарын ауа райы, жол кептелістері және тұтынушылар орналасқан жері туралы деректеріне негізделіп оңтайластырады. Бұл жеткізу уақытын азайтып, қызмет сапасын жақсартады.

7. **Әлеуметтік және мемлекеттік басқару** Қорытынды ретінде біз мемлекеттің дамуындағы әлеуметтік-экономикалық жағдайды талдауда, халықтың қандай қажеттіліктері бар екенін анықтауда және саясат әзірлеу үшін деректер ғылымын қолданатын қатерді атаймыз. Мұндай мәліметтер деректер базасынан алынуы мүмкін бір түрі ретінде жүргізіледі, мысалы: халық санағы, еңбек нарықтары, денсаулық сақтау және экология сияқты мәліметтер алу үшін мемлекет пайдалана алатын. "Ақылды қала" концепті бұл үрдістің бөлігі болып табылады және ол деректер арқылы қоғамдық көлікті, коммуналдық қызметтерді, қауіпсіздікті және қоршаған ортаны басқаруға назар аударады. Бұл өз кезегінде қала тұрғындарының өмір сапасын жақсартады және олар тең бөлінеді.

8. **Ғылым және зерттеу** Ғылым әлемінде деректер жаңа жаңалықтарды ашуға мүмкіндік береді. Астрономдар телескоптарды қолдана отырып, ғаламдағы жұлдыздардың қозғалысын зерттеу үшін деректер жинайды. Биологтар аурулардың пайда болуы және жаңа дәрілерді шығару әдістерімен геномдық деректер жиынтығын пайдаланады. Климатологияда климаттың өзгеруін модельдеу үшін үлкен деректер қолданылады. Мысалы, НАСА және ЕКА (ESA) сияқты агенттіктер климат өзгеруі жер бетіндегі өзгерістерді бақылап, табиғи апаттардың алдын алу үшін оның спутниктік деректерін қолданады.

Ақырында, деректер ғылымы кәсіп немесе құрал емес; әлемге жаңа тәсілдермен қараудың жолы болып табылады. Деректерді түсініп, оны қолдана алатын адамдар мен институттар тек әрекет етіп қана қоймай, оларды өзгерте алады. Осылайша, деректер ғылымының барлық аспектілері (мәселелері) болашақтың негізі бола алады.

Әдебиеттер тізімі:

1. Университеттің оқу бағдарламасы: Data Science. Наука о данных с нуля / Джоэл Грас. — [Электронный ресурс]. — Режим доступа: https://www.ursmu.ru/userfiles/media/default/sveden/education/1.2.1_iiimo-21_mm_kaf.inf.pdf
2. Sarker, I. H. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. — *SN Computer Science*, 2021, Vol. 2, Article 377. — DOI: 10.1007/s42979-021-00765-8.
3. Raj Kamal, Preeti Saxena. Big Data Analytics: Introduction to Hadoop, Spark, and Machine Learning. — Tata McGraw Hill Education, 2018. — [Электронный ресурс]. — Режим доступа: <https://jcer.in/jcer-docs/E-Learning/Digital%20Library%20/E-Books/Big%20Data%20Analytics%20by%20RaJ%20Kamal,%20Preethi%20Saxena,TMH.pdf>
4. IBM. Что такое наука о данных? — [Электронный ресурс]. — Режим доступа: <https://www.coursera.org/learn/what-is-datascience>