

Дәріс 1. Деректер туралы ғылымына кіріспе

Деректер туралы ғылымның негіздеріне кіріспе дәрісіне қош келдіңіз. Бұл дәрісте біз деректер ғылымының қандай екенін, оның қоғамдағы рөлін және қазіргі заманның ең маңызды кәсіби дағдыларының бірі болып табылатын деректерді өңдеу қабілетін қарастырамыз.

Оқытушы: Сарсембаева Т.С.



Дәрістің мақсаты

Негізгі ұғымдар

Деректер туралы ғылымның (Data Science) негізгі ұғымдарымен таныстыру

Қоғамдағы рөлі

Оның қазіргі қоғамдағы рөлімен және пішінімен таныстыру

Салалар және құралдар

Саланың негізгі салаларын, құралдарын және ақпараттың қалай жиналатынын түсіндіру

Кәсіби дағдылар

Деректерді өңдеу қабілетінің ХХІ ғасырдағы ең маңызды кәсіби дағдылардың бірі екенін көрсету

Негізгі сұрақтар

1 Деректер туралы ғылым дегеніміз не?

Data Science-тің анықтамасы және оның мәні

3 Деректерді зерттеушілер деген кімдер?

Data Scientists-тің міндеттері және құзыреттіліктері

2 Деректердің рөлі қандай?

Қазіргі қоғамда деректердің маңызы және қолданылуы

4 Мысалдар мен бағыттар

Деректер ғылымының практикалық қолданылуы

Қысқаша мазмұн: Деректер үстемдігі дәуірі

Біз деректер үстемдігі дәуірінде өмір сүріп жатырмыз. Осы заманғы қоғамдағы бірде-бір секунд әртүрлі құрылғылардан деректер жинап, сақтап, өңдемей өтпейді.



Осы шексіз деректер теңізінде адамзаттың міндеті – ақпараттың барлығын мәніне қарай саралап, оны шынайы білім мен шешімдерге айналдыру. Деректер ғылымы — бұл математика, статистика, бағдарламалау және салалық білімнің тоғысқан жері. Ол деректерден пайдалы үлгілерді, болжауларды және ақпараттандырылған шешімдерді алуға көмектеседі. Бұл ғылымның негізгі мақсаты – жай ғана сандар мен мәтіндер жинағын емес, деректерді мағыналы әрі пайдалы ақпаратқа айналдыру.

Кіріспе: Деректер үстемдігі дәуірінде өмір

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 XXI 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.

01	02	03
Деректерді жинау	Деректерді тазалау және өңдеу	Талдау және модельдеу
Әртүрлі көздерден деректер жинау: интернет беттері, сенсорлар және әлеуметтік желілер немесе сауалнамалар	Қажетсіз немесе қателерді ақпараттан тазалау, форматтарды нормалау	Статистикалық және машиналық оқыту алгоритмдерін қолдану арқылы қорытынды алу
04	05	
Нәтижелерді интерпретациялау	Шешімдер қабылдау	
Нәтижелерді басқаларға түсінікті түрде жасау, теңдеулер немесе диаграммаларды қолдану	Талдауды әрекетке көшіру кезеңі	

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. Деректер ғылымы жұмыс процесінің өз кезеңдері бар және әрбір кезең маңызды. 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.



Бизнесте

Нарықтағы сұранысты болжауға, бәсекелестерімнің не істеп жатқанын және тұтынушылардың қалай әрекет ететіні туралы түсінуге көмектеседі



Білім беруде

Студенттерді жетістіктерін бағалау, оқу процесін дербестендіру және сапалы білім алу үшін деректерді пайдалануды үйретеміз



Денсаулық сақтауда

Аурудың ерте сатыларын анықтауға, эпидемиялардың қалай әрі қаншалықты тез таралатынын болжауға және емдеу әдістерінің тиімділігін түсінуге көмектеседі



Ғылымда

Күрделі табиғи немесе әлеуметтік құбылыстардағы үлгілер мен құрылымдарды ашуға көмектеседі

Деректер технологиялық фетиш ретінде қолданылады және стратегиялық және қоғамдық маңызға ие. Деректер қазіргі өмірдің әрбір бұрышына және қуыс-қуыс түстеніп отыр. 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 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

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.

Бизнес және маркетинг

Қазіргі әлемде бизнес деректерге тең келеді. Компаниялар деректер ғылымын өздерінің тұтынушылары, нарық трендтерін болжау және табысты арттыру үшін қолданады. Amazon, Netflix немесе Spotify сияқты үлкен компаниялар деректер ғылымын қолданушылардың әрекеттеріне негізделген ұсыну жүйелерін құрастыру үшін пайдаланады. Деректер ғылымы компанияларға өз жарнамаларын неғұрлым нақты мақсатқа бағыттауда, аз шығынмен көмек көрсетеді.

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. Жасанды интеллект алгоритмдері рентген және MRI суреттерін қарап, ерте онкологиялық немесе жүрек ауруларына байланысты алғашқы сигналдарды автоматты түрде анықтай алады.

Білім беру

Білім беруде деректер ғылымы студенттердің жетістіктерін қадағалау және білім беру жүйесінің сапасын жаңғыртуға негізделген оқу процесін ұсынады. Онлайн платформалар (Coursera, edX, Khan Academy) студенттердің оқу әрекеттерін қадағалап, олар қиындық келтіретін тақырыптарды анықтап, қосымша ресурстар ұсынады.

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.

Өнеркәсіп және өндіріс

Өндіріс 4.0 дәуірінде өндіріс процесі тек деректер арқылы басқарылады. Мыңдаған датчиктердің ішінде машиналардың күйі, температура, қысым немесе өнім сапасы тұрақты тексеріледі. Инженерлер деректерді қарап отырып, ақауларды ерте ұстап және өндіріс тоқтағанға дейін жөндеу жұмыстарын жүргізе алады.

Көлік және логистика

Көлік және логистика саласында деректер ғылымы оңтайлы жолдарды құру және уақыт пен шығынды үнемдеуде қолданылады. Uber, Bolt, Yandex Go сияқты платформалар жолаушылар мен жүргізушілерді мүмкін болатын қысқа сәйкес келетін нақты уақыт деректеріне сәйкес байланыстырады.

Әлеуметтік және мемлекеттік басқару

Мемлекеттің дамуындағы әлеуметтік-экономикалық жағдайды талдауда, халықтың қандай қажеттіліктері бар екенін анықтауда және саясат әзірлеу үшін деректер ғылымын қолданады. "Ақылды қала" концепті деректер арқылы қоғамдық көлікті, коммуналдық қызметтерді, қауіпсіздікті және қоршаған ортаны басқаруға назар аударады.

Ғылым және зерттеу

Ғылым әлемінде деректер жаңа жаңалықтарды ашуға мүмкіндік береді. Астрономдар телескоптарды қолдана отырып, ғаламдағы жұлдыздардың қозғалысын зерттеу үшін деректер жинайды. Биологтар геномдық деректер жиынтығын пайдаланады. Климатологияда климаттың өзгеруін модельдеу үшін үлкен деректер қолданылады.

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.

Дәрісті меңгеруге арналған бақылау сұрақтары

1 Деректер туралы ғылым дегеніміз не?

Деректер туралы ғылым (Data Science) дегеніміз не және ол қандай пәндердің тоғысында дамыды?

2 Жұмыс процесінің кезеңдері

Деректер ғылымының жұмыс процесі қандай негізгі 5 кезеңнен тұрады?

3 Деректердің маңызы

Неліктен деректерді "жаңа дәуірдің мұнайы" деп атайды?

4 Деректерді зерттеушінің рөлі

Деректерді зерттеуші (Data Scientist) деген кім және оның негізгі міндеттері қандай?

5 Практикалық қолданылуы

Денсаулық сақтау, бизнес немесе қаржы салаларында деректер ғылымының қолданылуына 2-3 мысал келтіріңіз

Әдебиеттер тізімі:

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_iimo-21_mm_kaf.inf.pdf

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.

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>

IBM. Что такое наука о данных? — [Электронный ресурс]. — Режим доступа: <https://www.coursera.org/learn/what-is-datascience>