

№1 Дәріс. Data Science-тың негіздері және оның қолдану салалары

Дәрістің мақсаты: студенттерге Data Science (деректер ғылымы) ұғымының мәнін, оның даму тарихы мен негізгі бағыттарын, әдістері мен кезеңдерін түсіндіру, сондай-ақ оның түрлі салалардағы (білім, медицина, бизнес, қаржы, экология және мемлекеттік басқару) қолданылу ерекшеліктерін талдау.

Дәрісте қарастырылатын сұрақтар.

1. Data Science ұғымы және оның қалыптасу тарихы.
2. Data Science-тың негізгі компоненттері мен тіректері.
3. Деректер ғылымының негізгі кезеңдері мен әдістері.
4. Data Science және Big Data арасындағы байланыс.
5. Data Scientist маманының рөлі мен дағдылары.
6. Data Science-тың қолдану салалары.
7. Қазақстандағы Data Science дамуы мен мемлекеттік бастамалар.
8. Data Science саласындағы этикалық және құқықтық мәселелер.

Кіріспе

Data Science немесе деректер ғылымы – қазіргі замандағы ең қарқынды дамып келе жатқан ғылыми және қолданбалы бағыттардың бірі. Бұл сала үлкен көлемдегі деректермен (Big Data) жұмыс істеу, оларды жинақтау, өңдеу, талдау, болжау және шешім қабылдау үдерістерін қамтиды. Data Science тек ақпараттық технологиялармен шектелмей, статистика, математика, жасанды интеллект, экономика және қоғамтану сияқты көптеген салалармен тығыз байланыста дамып келеді. Оның басты мақсаты – деректерден нақты әрі пайдалы білім алу және сол білім арқылы өндірістік, экономикалық және әлеуметтік мәселелерді тиімді шешу.

Data Science-тың негізгі бағыттары мен құрамдас бөліктері.

Data Science үш негізгі тірекке сүйенеді:

1. Математика және статистика – деректердің заңдылықтары мен үлгілерін анықтау, ықтималдықтарды есептеу және нәтижелерді болжауға мүмкіндік береді.
2. Бағдарламалау – Python, R, SQL сияқты тілдер арқылы деректерді өңдеу, талдау және визуализациялау жүзеге асады.
3. Салалық білім – нақты қолдану саласындағы (мысалы, медицина, экология, қаржы) деректердің ерекшеліктерін түсінуге көмектеседі.

Деректер ғылымының кезеңдері.

1. Деректерді жинау – әртүрлі көздерден (интернет, мемлекеттік базалар, әлеуметтік желілер, IoT құрылғылары) мәліметтерді алу.
2. Деректерді тазалау және алдын ала өңдеу – артық немесе қате ақпаратты жою, деректерді құрылымдау.

3. Деректерді талдау – статистикалық және аналитикалық әдістер арқылы заңдылықтарды анықтау.
4. Модель құру және машиналық оқыту – деректер негізінде болжау немесе шешім қабылдау жүйелерін жасау.
5. Нәтижені визуализациялау және интерпретациялау – алынған нәтижелерді диаграммалар мен графиктер түрінде көрсету.

Data Science және Big Data.

Big Data – көлемі, жылдамдығы және әртүрлілігімен ерекшеленетін деректер жиынтығы. Data Science осы деректерден мән мен үлгі табу үшін ғылыми және алгоритмдік әдістерді қолданады. Мысалы, банктер клиенттердің транзакцияларын талдап, алаяқтық әрекеттерді анықтай алады, ал логистикада жеткізу жолдарын оңтайландыруға мүмкіндік туады.

Data Scientist маманы және оның дағдылары.

Data Scientist – деректерді жинау, өңдеу, талдау және машиналық оқыту модельдерін жасауға маманданған сарапшы. Оның негізгі дағдыларына Python, R, SQL тілдерін білу, Pandas, NumPy, Matplotlib, Scikit-learn кітапханаларымен жұмыс істеу, статистика мен ықтималдық теориясын меңгеру, визуализация құралдарын қолдану және аналитикалық ойлау қабілеттері жатады.

Data Science-тың қолдану салалары.

1. Бизнес пен маркетингте – тұтынушылардың мінез-құлқын талдау, сұранысты болжау, жарнаманы оңтайландыру.
2. Қаржы саласында – несиелік тәуекелдерді анықтау және қаржылық алаяқтықтың алдын алу.
3. Медицинада – ауруларды ерте анықтау, диагностика жүргізу, жаңа препараттарды зерттеу.
4. Білім беру жүйесінде – оқу нәтижелерін талдау, студенттердің жеке оқу траекториясын құру.
5. Мемлекеттік басқаруда – халықтың әлеуметтік деректерін талдау, тиімді шешім қабылдау.
6. Экология мен энергетикада – табиғи ресурстарды бақылау және энергия тұтынуды оңтайландыру.

Қазақстандағы Data Science дамуы

Қазақстанда “Цифрлық Қазақстан” бағдарламасы Data Science бағытындағы негізгі қозғаушы күш болып табылады. “Smart Astana” жобасында қала инфрақұрылымын, көлік қозғалысын және ауа сапасын бақылау жүйелері Data Science әдістерімен басқарылады. “Kazhydromet” пен “EcoAir” ұйымдары да экологиялық деректерді талдауда деректер ғылымын қолданады. Университеттер ішінде Өл-Фараби атындағы ҚазҰУ, Astana IT University, ҚБТУ және Satbayev University Data Science мамандарын даярлауда.

Этикалық және құқықтық аспектілер

Деректермен жұмыс кезінде жеке ақпараттың қауіпсіздігін сақтау өте маңызды. Қазақстан Республикасының “Дербес деректер және оларды қорғау туралы” заңы (2013, толықтырулармен 2024) осы бағыттағы негізгі құқықтық

негізді қамтамасыз етеді. Data Scientist маманы тек техникалық емес, этикалық жауапкершілік тұрғысынан да сауатты болуы тиіс.

Қорытынды/

Data Science – XXI ғасырдың ең маңызды және болашағы зор салаларының бірі. Ол экономиканы, ғылымды және қоғамды деректер негізінде дамытуға бағытталған. Қазақстанда бұл бағыттағы мамандардың қажеттілігі жыл сайын артып келеді, себебі Data Science – цифрлық экономиканың негізі мен инновациялық шешімдердің басты құралы.

Бақылау сұрақтары

1. Data Science дегеніміз не және ол қандай мақсаттарды көздейді?
2. Data Science-тың негізгі үш тірегін атаңыз.
3. Деректер ғылымының кезеңдері қандай?
4. Big Data және Data Science арасындағы байланыс неде?
5. Data Scientist маманы қандай дағдыларға ие болуы тиіс?
6. Data Science медицина және білім саласында қалай қолданылады?
7. Қазақстанда Data Science қандай жобаларда іске асырылып жатыр?
8. Data Science саласындағы этикалық мәселелер қандай?
9. Деректерді визуализациялау не үшін қажет?
10. Неліктен Data Science болашақтың мамандығы болып саналады?

Пайдаланылған әдебиеттер

1. Қазақстан Республикасы Цифрлық даму, инновациялар және аэроғарыш өнеркәсібі министрлігі. *«Цифрлық Қазақстан» мемлекеттік бағдарламасы*, 2024.
2. Әл-Фараби атындағы Қазақ ұлттық университеті. *Ақпараттық жүйелер және деректер талдауы пәнінің оқу-әдістемелік құралы*, Алматы, 2023.
3. Бекжанова А.Қ., Төлегенова А.М. *Data Science негіздері: оқу құралы*, Астана: Astana IT University, 2022.
4. Құдайбергенов Е.Т. *Big Data технологиялары және олардың Қазақстан экономикасында қолданылуы*, Алматы, 2021.
5. Satbayev University. *Жасанды интеллект және деректер талдауы курстарының жинағы*, Алматы, 2024.
6. Қазақстан Республикасының *«Дербес деректер және оларды қорғау туралы» заңы*, 2013 (өзгерістермен 2024).
7. *digitalkz.kz* — Қазақстанның цифрлық даму платформасы.
8. *data.gov.kz* — Қазақстанның ашық деректер порталы.