Дәріс 4

Үлкен деректерді сақтау және өңдеу технологиясы

Big Data технологиясының қолданылу аясы өте кең, ол кез келген түрдегі деректерді талдауға арналған әмбебап құралдар жиынтығы болып табылады. Оның ішінде көптеген салалар бар, соның ішіндегі негізгілері:

– Маркетинг

Деректерді интеллектуалды талдау технологиялары қолданылған ең алғашқы саланың бірі маркетинг саласы болып есептеледі. Big Data әдістерінің дамуы басталған тапсырма сатып алушылар себетін (корзинасын) талдау деп аталады. Берілген тапсырманың міндеті – сатып алушылар бірге сатып алуға ұмтылатын тауарларды анықтау. Сатып алушы корзинасын білу жарнамалық компанияларды өткізуге, сатып алушыларға жеке ұсыныстар қалыптастыруға, тауарлар қорын құру және оларды сауда залдарында орналастыру стратегиясын әзірлеу үшін керек. Сонымен қатар маркетингте қандай да бір тауарларды сәтті өткізу үшін оның мақсатты аудиториясын анықтау; кәсіпорындарға тауар қорларын құру жөнінде шешім қабылдауға көмектесетін уақытша үлгілерді зерттеу; кәсіпорындарға белгілі бір мінез-құлқы бар, әртүрлі санаттағы клиенттердің қажеттіліктер сипатын білуге мүмкіндік беретін болжау модельдерін құру; құнды клиентті жоғалтуды алдын алу және клиенттің мінезін талдау арқылы оның кету сәтін алдын ала анықтауға мүмкіндік беретін, клиенттің адалдығын болжау сияқты тапсырмалар шешіледі.

– Өнеркәсіп

Бұл саладағы өте маңызды бағаттардың бірі болып мониторинг және сапаны бақылау болып табылады, мұнда талдау құралдары көмегімен жабдықтардың істен шығуын болжауға, аққаулардың пайда болуы, жөндеу жұмыстарын жоспарлауға мүмкін болады. Белгілі бір сипаттамалардың танымалдығын болжау және қандай сипаттамалар бірге тапсырысқа берілетінің білу өндірісті оңтайландыруға, оны нақты тұтынушы қажеттіліктеріне бағыттауға көмектеседі.

– Медицина

Медицинада деректерді талдау сондай-ақ өте сәтті қолданылуда. Оған мысал ретінде тексеру нәтижелерінің талдауы, диагностика, емдеу әдістерінің тиімділігі мен дәрі-дәрмекті салыстыру, ауруларды талдау және олардың таралуы, жанама әсерлерді анықтау жатады. Big Data-тің ассоциативті ережелер және тізбекті үлгілер сияқты технологиялары дәрілерді қабылдау мен олардың жанама әсерлері арасындағы байланысты анықтауда сәтті қолданылады.

– Молекулалық генетика и гендік инженерия

Тәжірибелі деректерде заңдылықтарды анықтау тапсырмасы молекулалық генетика и гендік инженерияда өте өткір және онымен бірге нақты қойылған. Мұнда ол тірі ағзаның фенотиптік қасиеттерін бақылайтын, генетикалық кодтар

ретінде түсінілетін белгілерді анықтау ретінде тұжырымдалады. Мұндай кодтар жүздеген, мыңдаған және одан да көп байланысқан элементтерді қамти алады. Деректердің аналитикалық талдауының нәтижесі болып, сонымен қатар, генетик-ғалымдармен табылған адам ДНК-сының тізбегіндегі өзгерістер мен әртүрлі аурулардың даму қаупі арасындағы тәуелділік болып табылады.

– Қолданбалы химия

Big Data әдістері қолданбалы химия саласында да қолданыс табады. Мұнда олардың қасиеттерін анықтайтын, қандай да бір байланыстардың химиялық құрылым ерекшелігін анықтау туралы жиі сұрақ туындайды. Әсіресе мұндай тапсырмалар, сипаттамасы жүздеген және мыңдаған құрылымдық элементтер мен олардың байланысын қамтитын күрделі химиялық байланыстарды талдау кезінде өзекті болады.

– Қылмыспен күрес

Қауіпсіздікті қамтамасыз етуде Big Data құралдары жақында ғана қолданыла бастады, бірақ қазіргі уақытта бұл салада деректердің интеллектуалды талдауының тиімділігін растайтын тәжірибелі нәтижелер алынған. Швейцариялық ғалымдармен болашақ қақтығыстарды болжау мақсатында наразылық қызметін талдау жүйесі және әлемдегі хакерлер қызметі мен кибер қатерлерді бақылау жүйесі жасалған болатын. Соңғы жүйе кибер қатерлер мен басқа ақпаратты қауіпсіздік тәуекелін болжауға мүмкіндік береді. Сонымен қатар, Big Data әдістері несиелік карталармен болатын алаяқтарды анықтауға сәтті қолданылады.

– Басқа бағдарламалар:

– Қауіп қатерді талдау. Мысалы, төленген өтініштермен байланысты болатын факторлар тізбегін анықтау жолымен, сақтандырушылар міндеттемелері бойынша шығындарды азайтуы мүмкін. АҚШ-та үлкен сақтандыру компаниясы, некеде тұрған адамдардың өтініші бойынша төленген соммалар жалғыз басты адамдардың өтінішімен болған соммаларды асып түсетінін анықтаған жағдай белгілі. Компания бұл жаңа ақпаратқа отбасылы клиенттерге жеңілдіктер ұсыну саясатын қайтадан қарастыру тұрғысынан қарады.

– Метеорология. Ауа райын нейрондық желілер әдісімен болжау, оның ішінде өзінен ұйымдастырылатын Кохонен картасы қолданылады.

– Кадрлық саясат. Талдау құралдары қызметкерлерді басқару бөлімдеріне ең үздік кандидаттарды түйіндемелерін талдау негізінде таңдап алуға, қажет қызметке мінсіз қызметкерлердің сипаттамаларын модельдеуге көмектеседі.

Big Data құралдары дәстүрлі түрде қымбат бағдарламалық өнімге жатады. Сондықтан күні кешеге дейін бұл технологияның негізгі тұтынушылары банктар, қаржылық және сақтандырушы компаниялар, ірі сауда кәсіпорындары болған, ал Big Data қолданылуын талап ететін негізгі тапсырма болып несиелік және сақтандыру қатерін бағалау, маркетингтік саясат, тарифтік жоспарлар мен

клиентпен жұмыстағы басқа да принциптерді әзірлеу болған. Соңғы жылдары жағдай белгілі бір өзгерістерге ұшырады: бағдарламалық қамтамасыз ету нарығында салыстармалы арзан болатын Big Data құралдары және тіпті еркін таралатын жүйелер пайда болды, бұл осы технологияны шағын және орта бизнес кәсіпорындарына қол жетімді болатындай жасады.

Ақылы құралдар мен деректерді талдау жүйелері арасындағы көшбасшы болып SAS Institute (SAS Enterprise Miner), SPSS (SPSS, Clementine) және StatSoft (STATISTICA Data Miner) табылады. Жетерлікті белгілі болып Angoss (Angoss KnowledgeSTUDIO), IBM(IBM SPSS Modeler), Microsoft (Microsoft Analysis Services) және (Oracle) Oracle Big Data шешімдері саналады.

Еркін бағдарламалық қамсыздандыруды таңдауы сондай-ақ түрлілігімен ерекшеленеді. JHepWork, KNIME, Orange, RapidMiner сияқты әмбебап талдау құралдары бар болғандай, мамандырылған құралдар да болады, мысалы Carrot 2 – мәтіндік деректерді кластерлуге және іздеу сұраныстарының нәтижелері үшін арналған FrameWork, Chemicalize.org – қолданбалы химия саласындағы шешім, NLTK (Natural Language Toolkit) – табиғи тілді өңдеуге (natural language processing) арналған құрал.