

## МАШИНАЛЫҚ ОҚЫТУ, OPINION MINING ЖӘНЕ ОНЫҢ ҚАЗІРГІ ҚОҒАМДАҒЫ МАҢЫЗЫ

**Жақсылық Әділет Қасымжанұлы**

6B06120 - «Бағдарламалық инженерия» білім беру бағдарламасының 2 курс студенті

**Әкімжан Айбек Серікұлы**

6B06118- «Ақпараттық жүйелер және технологиялар»

білім беру бағдарламасының 4 курс студенті

М.Х.Дуллати атындағы Тараз университеті, Тараз қ., Қазақстан Республикасы

Ғылыми жетекшісі **Адилова Ақнұр Қалымбетқызы**

[ak.adilova@dulaty.kz](mailto:ak.adilova@dulaty.kz)

Машиналық оқыту (Machine Learning, ML) соңғы жылдары көптеген ғылыми және өндірістік салаларда кеңінен қолданылатын негізгі бағытқа айналды. Бұл сала ақпараттық технологиялар мен деректер ғылымының түйінді бөлігі болып табылады, адамның интеллектісіне ұқсас есептерді автоматты түрде шешу мүмкіндіктерін ұсынады. Машиналық оқыту негізінде идеялар мен принциптер осыдан бірнеше онжылдықтар бұрын ұсынылғанымен, соңғы жылдары оның дамуы жаңа биіктерге көтерілді. Деректердің көлемі мен түрі өсіп, техникалық мүмкіндіктер жақсарған сайын машиналық оқыту әдістері неғұрлым нақты әрі тиімді жұмыс істей алады. Бұл әдіс компьютерлердің өздігінен үйренуін, жаңа деректермен жұмыс істеуін және шешім қабылдау қабілетін дамытуды мақсат етеді. Машиналық оқыту жүйелері өздерін тәжірибеден үйренуге негізделген алгоритмдерді қолданғандықтан жаңа және күрделі міндеттерді шешуге пайдалану мүмкіндігі өте жоғары болып саналады.

Машиналық оқытудың ішінде ерекше орын алатын бағыттардың бірі ол жіктеу. Жіктеу - бұл деректерді белгілі бір категорияларға немесе сыныптарға бөлу процесі деп айта аламыз. Мысалы, мәтіндерге немесе суреттерге қатысты шешімдер қабылдау кезінде мәліметтер белгілі бір кластарға бөлінеді. Классификаторлар өздеріне тән бірнеше алгоритмдер мен техникаларды қолдана отырып, деректерді сараптап, нәтижесінде дұрыс шешім қабылдайды. Классификация көптеген қолданбалы салада қолданылады, мысалы, медициналық диагноздар, тұтынушы пікірлерін талдау, спам хаттарын анықтау және т.б жатады. Классификаторлар негізінде жұмыс істейтін жүйелер осы күндері адамдардың өмірін жеңілдететін түрлі қызметтерді ұсынады, олар өз кезегінде машиналық оқытудың дамуымен байланысты.

Жіктеуіштер - бұл мәліметтерді алдын ала белгіленген категорияларға немесе кластарға бөлу үшін қолданылатын машиналық оқыту әдістері. Жіктеуіштердің басты мақсаты - деректер жиынтығынан алынған үлгілерді немесе объектілерді сол деректерге негізделген белгілі бір сыныптарға дұрыс жатқызу. Өртүрлі жіктеуіштер өртүрлі тәсілдермен жұмыс істейді және олардың тиімділігі көптеген факторларға, соның ішінде деректердің ерекшеліктеріне, шешім қабылдау жылдамдығына және оқу кезеңіндегі күрделілігіне байланысты өзгеріп отырады.

Жіктеуіштің жұмыс істеу принциптері негізінен олардың қандай әдіс қолданатынына байланысты өзгереді. Бұл бөлімде сызықтық жіктеуіштер, сызықтық емес жіктеуіштер және ансамбль әдістері туралы қысқаша түсінік беріліп, олардың жұмыс принциптері қарастырылады.

Сызықтық жіктеуіштер - бұл деректерді сызықтық шекарамен бөлетін алгоритмдер. Бұл әдістер деректердің сызықтық түрде бөлінетін жағдайларында жақсы нәтижелер береді, яғни екі класты нақты бөлетін түзу немесе гипер парапараметрді табу үшін қолдануға болады.

Логистикалық регрессия сызықтық жіктеуіштің ең танымал мысалдарының бірі

болып саналады. Бұл әдіс негізінен екі немесе одан көп кластарды сызықтық шекара арқылы бөлуді мақсат етеді. Логистикалық регрессия арқылы алынған модель деректерді белгілі бір ықтималдықпен кластарға жатқызуға болады. Мысалы, электрондық пошта хатын спам немесе спам емес деп бөлу үшін қолдануға болады. Логистикалық регрессияның басты ерекшелігі ол деректердің сызықтық шекарамен бөлінетінін болжайды, яғни деректерді екі жақты ықтималдықтармен сипаттайды.

Алгоритмнің жұмыс принципін айта кетсек: алдымен әрбір дерек нүктесіне сәйкес сызықтық модель құрылады, содан кейін модельдің параметрлері (салмақтар мен ығысу) оптимизацияланады. Оңтайландыру процесі жиі градиенттік түсу әдісімен жүзеге асырылады, ол модельдің параметрлерін минимизациялау үшін пайдаланылады. Логистикалық регрессия сызықтық жіктеуіштер болып табылады, өйткені ол деректерді сызықтық шекара арқылы бөледі.

Шешім ағаштары - сызықтық емес жіктеуіштердің бір түрі болып табылады. Бұл әдіс деректерді бөлшектеу арқылы шешім қабылдайды. Әрбір түбірден немесе тармақтан белгілі бір шартқа сәйкес бөлінген деректер пайда болады. Шешім ағашының әрбір тармағы бір сипаттамадан тұрады, ал ұшындағы жапырақтар әр түрлі кластарды қарастырады.

Шешім ағаштарының басты ерекшелігі - олар деректерді түсіну мен түсіндіруді жеңілдетеді, өйткені олар ағаш түрінде ұсынылған шешімдер жиынтығынан құралады. Сонымен қатар, шешім ағаштары деректердің күрделі үлгілерін де талдай алады, бірақ олар жиі жоғары өзгергіштікпен сипатталады.

Қолдау векторлық машина (SVM) - сызықтық емес жіктеуіштердің тағы бір мысалы. SVM алгоритмі деректерді жоғары өлшемді кеңістікте сызықтық түрде бөледі, бірақ ол кеңістікті дұрыс бөлетін ең жақсы гипер парапараметрді табуға тырысады. Бұл әдіс тек екі класты жіктеуіштер үшін ғана емес, көп кластар үшін де қолдануға болады.

SVM алгоритмінің басты принципі - ол екі сыныпты ең үлкен арақашықтықпен бөлуді мақсат етеді, яғни деректердің ең жақсы шекарасын таба алады. Бұл әдіс әсіресе деректердің сызықтық емес түрде бөлінуі жағдайында өте тиімді болып табылады. Сонымен қатар, ядролық әдістерді қолдану арқылы SVM алгоритмі деректерді жоғары өлшемді кеңістікте бөлуге мүмкіндік бере алады, бұл сызықтық емес кластарды оңай ажыратуға көмектеседі.

Бұған қоса, машиналық оқытудың қазіргі кезеңдегі ең қызықты және көп зерттелетін бағыттарының бірі - пікірлерді талдау (opinion mining). Пікірлерді талдау - мәтіндер мен әлеуметтік желілердегі ақпараттарды зерттей отырып, олардан пайдаланушылардың эмоциялары мен пікірлерін автоматты түрде анықтау процесі. Бұл сала әсіресе бизнес, маркетинг, саясат және қоғамдық пікірді зерттеу салаларында үлкен рөл атқарады. Әлеуметтік желілерде және интернет алаңдарында күн сайын миллиондаған пікірлер, жазбалар мен жарияланып жатыр. Бұл пікірлердің арасында тұтынушылардың өнімдер мен қызметтерге деген көзқарастарын, көңіл-күйлерін және ойларын білдіретін ақпарат бар. Opinion mining, яғни пікірлерді талдау жүйелері осы ақпаратты өңдеуге, талдауға және маңызды нәтижелерге қол жеткізуге мүмкіндік береді.

Пікірлерді талдаудың маңыздылығы тек бизнес мақсаттарымен ғана шектелмейді, сонымен қатар тұтынушы пікірлерін зерттеу арқылы қоғамдағы трендтерді анықтауға болады. Мысалы, белгілі бір өнім немесе қызметке қатысты пікірлерді талдай отырып, тұтынушылардың қандай проблемалар мен сұраныстар бар екенін анықтауға болады. Бұл мәліметтер маркетинг стратегияларын, өнімдерді жақсарту, тұтынушы тәжірибесін жақсарту үшін өте маңызды болып саналады. Сонымен қатар, қоғамдық пікірді зерттеу мен қоғамдық көзқарасты бағалау саяси шешімдер мен әлеуметтік жағдайларды бақылауда да тиімді. Бұл тақырыптан түйгеніміз машиналық оқыту, жіктеуіштер және пікірлерді талдау кезінде қазіргі ақпараттық қоғамда маңызды рөл атқаратын технологиялық бағыттар болып табылғандықтан олар көптеген салаларда автоматтандыру мен тиімділікті арттыру үшін қолданылуда. Жіктеуіштер мәліметтерді тиімді талдауға және нақты шешімдер қабылдауға көмектеседі, ал пікірлерді талдау тұтынушылардың көзқарастарын зерттеп, оларды бизнес

стратегиялары мен шешімдер қабылдауда пайдалану мүмкіндігін ұсынады. Машиналық оқыту мен пікірлерді талдау әдістері кеңінен дамып, кәсіпорындар мен ұйымдарға түрлі қызмет түрлерін оңтайландыруға, маркетингтік стратегияларын жетілдіруге, сондай-ақ қоғамдық пікір мен тұтынушы көңіл-күйін тиімді түрде бақылауға мүмкіндік береді.

Машиналық оқыту әдістерін қолдана отырып, пікірлерді талдау (opinion mining) мәселесін зерттеп және оның қоғамдағы маңызын ашамыз. Бұл зерттеуімізде пікірлерді талдау әдісінің шектеулерін қарастырып мүмкіндіктерін ұсынамыз. Жұмыстың міндеті машиналық оқыту әдістерін зерттей отырып олардың пікірлерді талдаудағы рөлін түсіндіреміз және пікірлерді талдаудағы ерекшелігін талқылап қолдану әдістерін қолданамыз.

Қазіргі уақытта машиналық оқыту (ML) информатика мен инженерия саласындағы қызықты зерттеу саласы болып табылады. Бұл жасанды интеллекттің ішкі жиынтығы болып саналады, өйткені ол үлгілерден мағыналы үлгілерді алуға мүмкіндік береді және бұл адамның интеллектінің мүмкіндігі. Қайталанатын және нақты анықталған тапсырмаларды орындайтын компьютерге деген сұраныс айқын: компьютерлер берілген тапсырманы дәйекті және қажымай орындайтын болады, бірақ бұл тапсырмаларды адам үшін орындау қиынға соғады. Соңғы жылдары машиналар үшін өте күрделі деп саналатын тапсырмаларды үйрену және тіпті меңгеру қабілетін көрсетті, бұл машиналық оқыту алгоритмдері анықтау және шешім қабылдауды қолдау жүйелерінің ықтимал пайдалы элементтері екенін көрсетті. Тағы бір қызықты жайтты ескерсек, кейбір жағдайларда компьютерлер адамның қабылдауынан тыс заңдылықтарды байқай алатын сияқты. Бұл жаңалық әртүрлі салалардағы машиналық оқыту мамандығына деген қызығушылықтың айтарлықтай және артуына әкелді. Жоғары деңгейде машиналық оқыту - бұл компьютерлік жүйеге деректерді беру кезінде дәл болжам жасауды үйрету процесі. Бұл болжамдар фотосуреттегі жемістердің бір бөлігі банан немесе алмама, жоқ па, өзін-өзі басқаратын көліктің алдында жолды кесіп өтіп бара жатқан адамдарды байқай ма, сөйлемдегі "кітап" сөзін қолдану қағаз мұқабалы кітапқа немесе қонақүйді брондауға қатысты ма, электрондық пошта спамба, әлде бейнеге субтитрлер жасау үшін сөйлеуді дәл танума деген сұраққа жауап болуы мүмкін.

Дәстүрлі компьютерлік бағдарламалық жасақтама мен машиналық оқыту тәсілінің басты айырмашылығы - адам әзірлеушісі жүйеге банан мен алманың айырмашылығын қалай ажыратуға болатынын көрсететін кодтар жазбаған. Оның орнына, машиналық оқыту моделі үлкен көлемдегі деректерді пайдалана отырып, жемістерді сенімді түрде қалай ажыратуға болатынын үйретті, бұл жағдайда олардың үлкен саны болуы мүмкін банан немесе алма бар деп белгіленген суреттер. Негізінен, машиналық оқытудың негізгі бағыты тәжірибе арқылы автоматты түрде жетілдірілетін компьютерлік бағдарламаларды қалай құруға болады деген сұрақпен айналысады. Сонымен, сіздің жауабыңыз сіздің деректеріңізде шын мәнінде, машиналық оқыту машиналарды тәжірибе арқылы жетілдіруге мүмкіндік беретін статистикалық әдістерді қолданатын жасанды интеллекттің ішкі жиынтығы ретінде қарастыруға болады. Бұл компьютерлік жүйеге белгілі бір тапсырманы орындау үшін шешім қабылдауға мүмкіндік береді. Бұл бағдарламалар немесе алгоритмдер жаңа деректерді бақылау арқылы уақыт өте келе үйреніп, жетілдіре алатындай етіп жасалған.

Машиналық оқытудың мақсаты - мәліметтерден мағынаны алу. Осылайша, деректер машиналық оқытудың құлпын ашудың кілті болып табылады. Машиналық оқыту (ML) деректері неғұрлым білікті болса, ML алгоритмі соғұрлым дәлірек болады.

Пікірлерді іздеу (сонымен бірге көңіл-күйді талдау) пайдалану процесін білдіреді, Табиғи тілді өңдеу (NLP) немесе машиналық оқыту (ML) әдістері мәтіндік мәліметтерден субъективті ақпаратты автоматты түрде анықтау, алу және талдау. Пікірлерді іздеудің мақсаты мен мәтіннің бір бөлігінде айтылған пікірді немесе оның оң, теріс немесе бейтарап екендігін анықтау. Қазіргі қоғамда пікір іздеу қоғамдық қабылдауды, тұтынушылардың мінез-құлқын және әлеуметтік тенденцияларды түсінуге ұмтылатын әртүрлі салалар мен ұйымдар үшін маңызды құралға айналды.

Қазіргі қоғамдағы пікір өндірудің рөлі. Интернеттің, әлеуметтік медианың және интернеттегі шолулардың пайда болуы адамдарға өнімдер мен қызметтерден бастап саяси оқиғалар мен әлеуметтік мәселелерге дейінгі көптеген тақырыптар бойынша өз пікірлерін білдіруді жеңілдетті. Дегенмен, бұл деректердің үлкен көлемі адамдарға уақытылы және тиімді түрде қолмен талдауды іс жүзінде мүмкін емес етеді. Пікірлерді іздеу бұл мәселені үлкен деректер жиынтығынан пікірлерді алу процесін автоматтандыру арқылы шешеді, бұл ұйымдарға қоғамдық көңіл-күй мен тенденциялар туралы құнды түсінік алуға мүмкіндік береді. Пікірлерді іздеудің маңыздылығының артуының басты себептерінің бірі - цифрлық платформаларда пайдаланушы жасаған мазмұнның жариялануы. Facebook, Twitter және Instagram сияқты әлеуметтік медиа сайттары, сондай-ақ Yelp, TripAdvisor және Amazon сияқты шолу платформалары пікірлердің кең қоймаларын жасады. Бұл платформалар пайдаланушыларға өз ойлары мен тәжірибелерін миллиондаған басқа пайдаланушылармен бөлісуге мүмкіндік береді, бұл бизнес үшін де, үкіметтер үшін де, зерттеушілер үшін де құнды ресурсқа жатады.

Қазіргі қоғамда пікірлерді іздеу бизнес үшін тұтынушылардың қанағаттанушылығын түсінуде немесе тұтынушылардың қалауын болжауда ғана пайдалы емес. Ол сондай-ақ саясатта, денсаулық сақтауда, ойын-сауықта және әлеуметтік зерттеулерде шешуші рөл атқарады. Мысалы, саяси партиялар кандидаттар немесе саяси ұсыныстар туралы қоғамдық пікірді бағалау үшін көңіл-күйді талдауды пайдалана алады. Медицина мамандары денсаулық сақтаудың жаңа тенденцияларын немесе белгілі бір медициналық мәселелерге қатысты қоғамдық алаңдаушылықты анықтау үшін онлайн талқылауларды талдай алады. Ойын-сауық индустриясы фильмдерді, музыканы немесе телешоуларды қабылдауды бақылау үшін пікір іздеуді қолдана алады. Әлеуметтанушылар қоғамның әртүрлі әлеуметтік және мәдени тақырыптарға деген көзқарасын түсіну үшін әлеуметтік желілердегі пікірталастардағы көңіл-күйді зерттей алады.

Пікірлерді іздеудің ең танымал қосымшаларының бірі тұтынушылардың көңіл-күйін талдау саласында. Кәсіпорындар тұтынушылардың қанағаттанушылығын өлшеу, ықтимал проблемаларды анықтау және соған сәйкес стратегияларын түзету үшін тұтынушылардың

пікірлерін және әлеуметтік желілердегі жазбаларын талдайды. Мысалы, Amazon-да тұтынушылардың өнім туралы пікірлерін алу арқылы компания олардың өнімін тұтынушылар жақсы қабылдағанын немесе шешуді қажет ететін жалпы шағымдардың бар-жоғын бағалай алады. Сонымен қатар, әлеуметтік желілердегі брендіне қатысты көңіл-күйді зерттей отырып, компаниялар өз компанияларының жалпы беделін бағалай алады. Пікірлерді іздеу сонымен қатар нарық тенденцияларын болжауға көмек бере алады. Клиенттердің жаңа өнімдерге немесе қызметтерге деген көзқарасын талдай отырып, кәсіпорындар сұранысты болжай алады және деректерге негізделген шешімдер қабылдайды. Теріс көңіл-күй өнімнің кемшіліктерін немесе қанағаттанбауын көрсетуі мүмкін, ал оң көңіл-күй өсу мен кеңею мүмкіндіктерін көрсете алады.

Саяси және қоғамдық пікірді талдау. Бүгінгі таңда саяси ғалымдар сайлаушылардың көңіл-күйін түсіну және сайлау науқанының стратегияларын түзету үшін қоғамдық пікірді зерттеуге көп көңіл бөледі. Твиттерді, Facebook жазбаларын және жаңалықтар мақалаларын талдай отырып, саяси партиялар кандидаттарға, саяси мәселелерге немесе үкіметтің әрекеттеріне қатысты жұртшылықтың көңіл-күйі туралы түсінік бере немесе ала алады. Мысалы, белгілі бір саясаткерге немесе саясаткерге қатысты елеулі жағымсыз сезімдер болса, бұл ақпарат саяси науқанның хабары мен тәсілін өзгерту үшін пайдалануы мүмкін. Қоғамдық пікірді зерттеу сонымен қатар адамдардың әртүрлі мәселелерге қалай қарайтынын нақты уақыт режимінде дәлірек бағалау арқылы қоғамдық пікірді анықтауға көмектеседі. Телефон арқылы сауалнама жүргізу сияқты дәстүрлі сауалнама әдістері баяу және шектеулі болуы мүмкін. Интернеттегі дискурсты талдай отырып, қоғамдық пікірді талдау саясаттанушыларға сайлаушылардың қалауын жақсырақ түсінуге көмектесу арқылы қоғамдық көңіл-күй туралы тезірек түсінік бере алады.

### **Қолданылған әдебиеттер тізімі**

1. М.Ф.Баймухамедов, А.М.Баймухамедова, С.Н.Боранбаев. Жасанды интеллект:казіргі заманты теория жоне тажірибе. Оқу кұралы, 1-бөлім/ - Алматы: Бастау ,2020. - 248 б.
2. П.А.Пылов. Основы работы с моделями машинного и глубокого обучения: учебное пособие / - Москва, Вологда: Инфра-Инженерия, 2023. - 256 с.
3. В.И. Горбаченко. Машинное обучение: учебное пособие / - М.: Ай Пи Ар Медиа, 2023. - 218 с.
4. С.В.Запечников. Основы интеллектуального анализа данных и машинного обучения. конспект лекций. учебное пособие / - Москва: Национальный исследовательский ядерный университет «МИФИ», 2022. - 136 с.
5. А.Ж.Асамбаев. Жасанды интеллект негіздері. - Алматы: Эверо, 2017. - 168 б.
6. L.Zhang & M.Zhang. (2016). Opinion Mining from Web Data. Springer.
7. Bing Liu. (2017). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.