

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН



ҚазҰТУ ХАБАРШЫСЫ

ВЕСТНИК КазННТУ

VESTNIK KazNRTU

№ 2 (138)

Главный редактор
И. К. Бейсембетов – ректор

Зам. главного редактора
Б.К. Кенжалиев – проректор по науке

Отв. секретарь
Н.Ф. Федосенко

Редакционная коллегия:

З.С. Абишева- акад. НАН РК, Л.Б. Атымтаева, Ж.Ж. Байгунчечков- акад. НАН РК, А.Б. Байбатша, А.О. Байконурова, В.И. Волчихин (Россия), К. Дребенштед (Германия), Г.Ж. Жолтаев, Г.Ж. Елигбаева, Р.М. Исаков, С.Е. Кудайбергенов, Б.У. Куспангалиев, С.Е. Кумеков, В.А. Луганов, С.С. Набойченко – член-корр. РАН, И.Г. Милев (Германия), С. Пежовник (Словения), Б.Р. Ракишев – акад. НАН РК, М.Б. Панфилов (Франция), Н.Т. Сайлаубеков, А.Р. Сейткулов, Фатхи Хабаши (Канада), Бражендра Мишра (США), Корби Андерсон (США), В.А. Гольцев (Россия), В. Ю. Коровин (Украина), М.Г. Мустафин (Россия), Фан Хуан (Швеция), Х.П. Цинке (Германия), Е.М. Шайхутдинов-акад. НАН РК, Т.А. Чепуштанова

Учредитель:

Казахский национальный исследовательский технический университет
имени К.И. Сатпаева

Регистрация:

Министерство культуры, информации и общественного согласия
Республики Казахстан № 951 – Ж “25” 11. 1999 г.

Основан в августе 1994 г. Выходит 6 раз в год

Адрес редакции:

г. Алматы, ул. Сатпаева, 22,
каб. 609, тел. 292-63-46
Nina. Fedorovna. 52 @ mail.ru

D.T. Kassymova, A.T. Akhmediyarova, A.S. Shayakhmetova, M. Turdalyuly
(Institute of Information and Computing Technologies of the Ministry of Education and Science of the Republic of Kazakhstan, Almaty, Kazakhstan. dikakassymova@gmail.com)

ANALYSIS OF METHODS TO IDENTIFY AND ELIMINATION CONTRADICTIONS IN BIG DATA

Abstract: The article analyzes the conflicts associated with the processing of big amounts of data and the methods used to resolve them. The challenge is to identify conflicting data between big data. The approaches to the identification and elimination of conflicting data are analyzed. The use of machine learning methods when working with conflicting data was considered.

Keywords: big data, contradiction, temporary contradiction, spatial contradiction.

Д.Т. Қасымова, А.Т. Ахмедиярова, А.С. Шаяхметова, М. Тұрдалыұлы
(ҚР БЖҒМ ҒК Ақпараттық және есептеуіш технологиялар институты,
Қазақстан, Алматы қаласы, e-mail: dikakassymova@gmail.com)

ҮЛКЕН ДЕРЕКТЕРДЕ КЕЗДЕСЕТІН ҚАЙШЫЛЫҚТАРДЫ АНЫҚТАУ МЕН ЖОЮҒА ҚОЛДАНЫЛАТЫН ӘДІСТЕРГЕ ТАЛДАУ

Түйіндемe: Бұл мақалада үлкен көлемді деректерді өңдеуде кездесетін қайшылықтарды анықтау және оларды жоюға қолданатын әдістерге талдама жасалады. Әр ортадан түскен үлкен деректер арасында қайшылықты деректерді анықтау ауқымды мәселелердің бірі. Қайшылықты деректерді анықтауға және жоюға арналған тәсілдемелер, ғылыми жұмыстарға талдау жасалынды. Машиналық оқыту әдістерінің қайшылықты деректермен жұмыс жасауда қолданылуына талдама жүргізілген.

Түйін сөздер: үлкен деректер, қайшылық, уақыттық қайшылық, кеңістіктік қайшылық

Kіріспе.

Қайшылықтар ұғымын көптеген ғалымдар кеңінен қарастырып әртүрлі анықтамалар берген. Солардың ішінде кейбір анықтамаларға тоқтала кетейік.

Қайшылық – диалектиканың басты категориясы ретінде – өзара бірлік және өзара сіңісу ахуалында тұрған нысан немесе жүйенің бірін-бірі жоққа шығаратын, бір-біріне қарама-қарсы келетін жақтарының өзара әрекеттесуі. (Уикипедия. Ашық энциклопедиясы). Қайшылық жалпыға бірдей сипатта, табиғатта да, ойлау жүйесінде де, санада да кездеседі. Белгілі бір мағынада қайшылық қарама-қарсылықтардың бір-бірін жоққа шығаруы (күресі) мен бірлігін, тұтастығын білдіреді.

Болмыстың қайшылыққа құрылғандығын алғаш байқаған грек ойшылы Гераклит дүниені үздіксіз қалыптасу деп қарағандықтан, оның әрбір кезеңін, қалпын екі жақта көрді: дүниенің негізі – от, ол бірде жанып, бірде өшіп тұрады. Оның пікірінше табиғат, бүкіл ғарыш – ұдайы өзгеріп, дамып отырады.

Неміс философы Гегель: “Заттардың бәрі өз ішінде қайшылықты келеді, “қайшылық – заттардың ақиқаты мен мәнін білдіреді”, – деді.

Логикалық қайшылық – бір уақытта А жағдайының бар болуы мен осы жағдайдың жоқ болуының қабылдануы. Жалған тепе-теңдік болатын логикалық тілде А және – А емес: $(A \& \neg A)$ тұжырымдамасымен белгіленеді.

Ақылды қала компоненттерін зерттей келе күнделікті өмірде көптеген қайшылықтарды кездестіретініміз анықталды. Қайшылықтар әр салада әртүрлі болады. Мысалы, түрлі деректер қорындағы ақпараттар, көлік жүйелерін есепке алу жүйелерінде және айыппұлды есепке алу кезіндегі, уақытқа байланысты ақпараттар және т.б.

Деректерді интеллектуалды талдау әдісі арқылы алынған білімнің сапасы мен өнімділігі тек қана әдістің құрылымы өнімділігіне байланысты емес, деректердің сапасы мен жарамдылығына да байланысты. Өкінішке орай, шу, жетіспейтін құндылықтар, сәйкессіздік және артық деректер сияқты теріс факторлар білім алу және білім алу үшін пайдаланылатын деректерге қатты әсер етеді. Сапасы нашар деректер білім сапасының нашарлауына әкелетіні белгілі [15]. Осылайша, деректерді алдынала өңдеу [16] негізгі және маңызды қадам болып табылады, оның міндеті - одан әрі деректерді өңдеу алгоритмдері үшін дұрыс және пайдалы деп санауға болатын соңғы деректер жиынтығын алу.

Деректер қорындағы қайшылықтарды негізгі үш топқа бөлуге болады [2]:

1. Атаулардағы қайшылықтар әртүрлі типтегі заттар үшін бірдей атауларды немесе бір объекті үшін бірнеше атауларды пайдаланудан тұрады;

2. Құрылымдық – ең жиі кездесетін қайшылықтардың бірі, ұқсас немесе сол объекті үшін құрылымы бойынша әртүрлі болып келетін модельдер, кілттер немесе амалдарды пайдалану;

3. Семантикалық қайшылықтар деректер немесе мағлұматтар мәні бойынша қарама-қайшы, мысалы, әлемді қабылдаудың әр түрлі жүйелері салдарынан олардың мәні қарама-қайшы болған кезде пайда болады.

Қайшылықты деректердің әр салада, әртүрлі деректер базаларында туындау факторлары 1-ші кестеде көрсетілген:

1-кесте. Қайшылық факторлары туындайтын салалар

№	Сала түрі	Қайшылық туындайтын факторлар
1	Халыққа жаппай қызмет көрсету орталығы	1. адамдардың тегі мен аты-жөндерін дұрыс жазбау; 2. туған күндерін қате жазу; 3. ЖСК дұрыс енгізу
2	«Сергек» жолды бақылау жүйесі	1. Көлік иесіне қате айыппұлдың келуі. Сергек камераларынан келген фото және видеофиксациялар.
3	Маршруттық автобустар (Автобус парктері)	1. Кесте бойынша уақыттан ауытқу; 2. Автобустардың істен шығуы; 3. Жолаушылар тасымалына байланысты автобустардың жүру аралығы;
4	Қала жолдарының орналасуы	1. Көшелердің орналасу қайшылығы
5	Банктік қызмет көрсету	1. Клиенттер туралы ақпараттардың қателігі
6	Теміржол көлігі	1.Бекет диспетчерлерінің ақпаратты қате беруі 2. Поездар қақтығысы 3. Поезд жолдарының істен шығуы

Үлкен деректердегі қайшылықтарды анықтау

Үлкен деректер анықтамасына алуан түрлілік, өзгергіштік және анықталғандық сияқты сипаттамалар кіреді. Түрлі типтегі көптеген деректер көзі, сондай-ақ үнемі өзгеріп отыратын деректер мәндерінің қиыспайтындығы, қателіктер мен бұрмалану ықтималдығының артуына, түрлі форматтағы деректерді үйлестіру мен олардың мәнін түсіндіруде қиындықтардың туындауына алып келеді. Бұл талдау түрлі сападағы және сенімділік деңгейіндегі деректердің болуын талап етеді. Сенімділікті арттыру үшін деректер көпсатылы тазарту мен түрлендіру процедураларынан өтеді. Үлкен деректердегі жаңа мағлұматтарды іздеу процесін бірнеше кезеңдерге бөлуге болады. Әрбір кезеңнің табысты болуы келесі факторларға байланысты болады: деректердің өміршеңдік кезеңінің алдыңғы кезеңінен «мұраланған» мәселелер саны; талдамалық платформа ұсынатын аспаптық өңдеу құралдарының жиыны [7];

Қазіргі уақытта үлкен деректерден алынатын ақпараттардың сапасына ерекше көңіл бөлінуде. Үлкен деректердегі қайшылықты деректерді анықтау алдын-ала өңдеу процессінде және деректер қорында орын алатындығы анық. Жоғарыда айтылып кеткен факторлар себебінен туындайтын қайшылықты деректерді анықтап және оларды жою әдістері қарастырылған.

Ең алдымен, қай ақпарат қайшылықты ақпарат екенін анықтап шешкеніміз жөн. Осыдан кейін бірнеше іс-әрекет нұсқаларын қарастырамыз:

1) бір-біріне қайшы келетін мәліметтерді тапқаннан кейін оларды жою. Ең қарапайым, бірақ ақылға қонымды әдіс емес.

2) қарама-қайшылықты деректерді дұрыстау (нұсқалардың бірі - әр қайшылықты оқиғаның болу ықтималдығын есептеу және ең ықтималдысын таңдау), ең сауатты әдіс.

Қайшылықтарды анықтау мен жою үшін негізінен, үлкен қол жұмысы мен алдын-ала дайындықты талап ететін, объектілер арасындағы байланыстардың аса егжей-тегжейлі сипаттамасын енгізу, нақты бір әрекеттерге тыйым салу немесе басқа іс-шаралар арқылы, атаулардағы қайшылықтар мен құрылымдық қайшылықтарды анықтауға үшін немесе қақтығыстарды алдын-алуға арналған түрлі тәсілдер мен аспаптық құралдар қолданылады [3].

Қайшылықтарды анықтау әдістерін қолдану, деректердегі қателерді іздеу және жөндеумен ғана шектелмейді, олардың көмегімен бұрмаланған деректерді де табуға болады. Мәтіндерді талдау алгоритмдері қайшылықтарды табу және түпнұсқа мәтіндердің жасанды құрылған көшірмелерін анықтау үшін пайдаланылады. Бірақ, тәжірибе көрсеткендей, мәтіндердің синтаксистік,

семантикалық және тақырыптық қасиеттерін пайдаланатын алгоритмдер толық емес құралған, жоғары есептеу күрделілігімен сипатталады, ал оларды қолдану экономикалық тұрғыдан тиімсіз [6].

Семантикалық қайшылықтарды табу үшін латентті-семантикалық талдау әдісін пайдаланған жөн. Бұл әдісті қайшылықтарды табу үшін қолдану тиімділігі дәлелденген. Семантикалық жақын ақпараттардағы қайшылықтарды шешу мүмкіндігі мен нұсқалары туралы шешімді Мамдани алгоритмін немесе нейронды желілерді пайдаланып, айқын емес қорытындылар әдісінің негізінде қабылдауға болады [3-8].

Уақытша және кеңістіктік қайшылықтар

Уақыт қатарының екі деректерінің ақпараты арасында қайшылық болған кезде уақыттық сәйкессіздік туындайды. Бұл жағдайда уақытқа сілтемесі бар деректер элементтері деректердің бір жиынтығына немесе түрлі деректер жиынтығына уақытша қайшылық келеді немесе қабаттасады. Атап айтқанда, нақты тұжырымдар деректер жиынтығының элементтеріне негізделген, мысалы, себеп-салдарға байланысты болғанда, мұндай сәйкессіздік өте маңызды болып табылады [7]. Ондай сәйкессіздіктер жекеленген немесе толық болуы мүмкін. 2-кестеде уақыттық қайшылық жағдайының туындауының мысалдары келтірілген.

2-кесте. Уақыттық қайшылықтар

Қайшылықты жағдайлар	
Ішінара	Екі сәйкес келмейтін оқиғаның уақыт интервалдарының ішінара жабылуы
Толық	Екі сәйкес келмейтін оқиғаның уақыт интервалы сәйкес келу немесе қанағаттандыруды тежеу
Ауытқымалы мән	Уақыт қатарларының деректерінің ауытқуы
Контекстік	Уақыт қатарының деректері берілген контексте аномалды жағдайы бар

Келісілмеген екі оқиғаның уақытша интервалдары ішінара жабылса уақытша ішінара сәйкессіздік пайда болады. Келісілмеген екі оқиғаның уақытша интервалдары сәйкес немесе тежеуді қанағаттандырса, онда толық сәйкессіздік пайда болады.

Кеңістіктік қайшылық геометриялық қасиеттермен және әр түрлі кеңістіктік қатынастарымен деректер жинағында кеңістіктік шектеулерді бұзған кезде пайда болады. Объектіні агрегаттау қандай да бір бірегей шектеулерден шыққан кезде бірнеше объектілердің арнайы өлшемі берілсе, онда деректерді интеграциялау жобасында жүзеге асуы мүмкін [6]. 3-кестеде кеңістіктік қайшылықтардың туындайтын жағдайларына мысалдар қарастырылған [8].

3-кесте. Кеңістіктік қайшылық

Қайшылықты жағдайлар	
Геометриялық орналасуы	Қайшылықты геометриялық тұрған жері бар кеңістіктік қайшылық
Геометриялық форма	Қайшылықты геометриялық формасы бар кеңістіктік қайшылық
Топологиялық	Топологиялық шектеулерді бұзушылық
Қашықтық	Қашықтық қасиеттерінің бұзылуы
Масштабтау	Әртүрлі масштабтағы түрлі геометриялық бейнелердің кеңістіктік ерекшелігі
Семантикалық шектеу	Шектеу тұтастығының мағыналық бұзылуы
Құрылымдық шектеу	Геометриялық примитивтердің шектелуінің құрылымдық тұтастығын бұзу
Интеграциялаудың туындауы	Әр түрлі көздерден бірдей кеңістіктік объектінің әртүрлі көріністері, нәтижесінде объектілерде ерекше геометриялық көрініс болуы керек деген шектеулер бұзылады.

Мәтіндік деректер көзінен алынған ақпараттарды агрегаттаудан да қайшылықтар туындауы мүмкін. Сонымен қатар, құрылымдалмаған мәтіндік деректер асимметриялы, антонимге, сәйкессіз мәндерге және сәйкессіздіктерге қатысты семантикалық және синтаксистік өлшемдермен берілетін қасиеттердің нысандарын көрсете алады. Қайшылықтар деректер терминдерінде, семантика терминдерінде және оларды құрылымдық ұсыну терминдері туындауы мүмкін [7]. Сондықтан деректердегі мәтіндік қайшылықтар мәселелерін талқылау кезінде тілдің семантикасы маңызды. Онтологияларға негізделген семантикалық

ережелерді құру мүмкіндігі қайшылықтарды анықтауда негізгі қадамы болуы мүмкін және олар қарастырылып отырған тұжырымдарды бағалауға және тексеруге мүмкіндік береді. Бұл мәселені шешудегі ең маңызды мәселелердің бірі қолда бар ақпаратты салмақтап, қайшылықты құндылықтар арасында ең жақсы деректер мәнін табуға болады.

Қайшылықтарды жою үшін көптеген тәсілдер әзірленді. [7] авторлары шешімдер мен стратегияларды қабылдау, стратегиялардағы конфликтерді жою әдістерін жіктеген.

Біріншісі деректер жиынында кездесетін қайшылықтардың нұсқасын деректерді айқындайды және мәніне немесе метадеректеріне байланысты түпкілікті мән ретінде тағайындайды. Соңғысы деректер жинынтығындағы мәндерді агрегаттау жолымен қайшылықтарды шешеді. Бұл стратегияның түрі мәндерге жатпайтын нәтижелерді таңдайды. Олардың арасындағы негізгі алғышарттардың бірі шешім қабылдау стратегиясы жағдайында тізбектің бақылануымен тұжырымдалады.

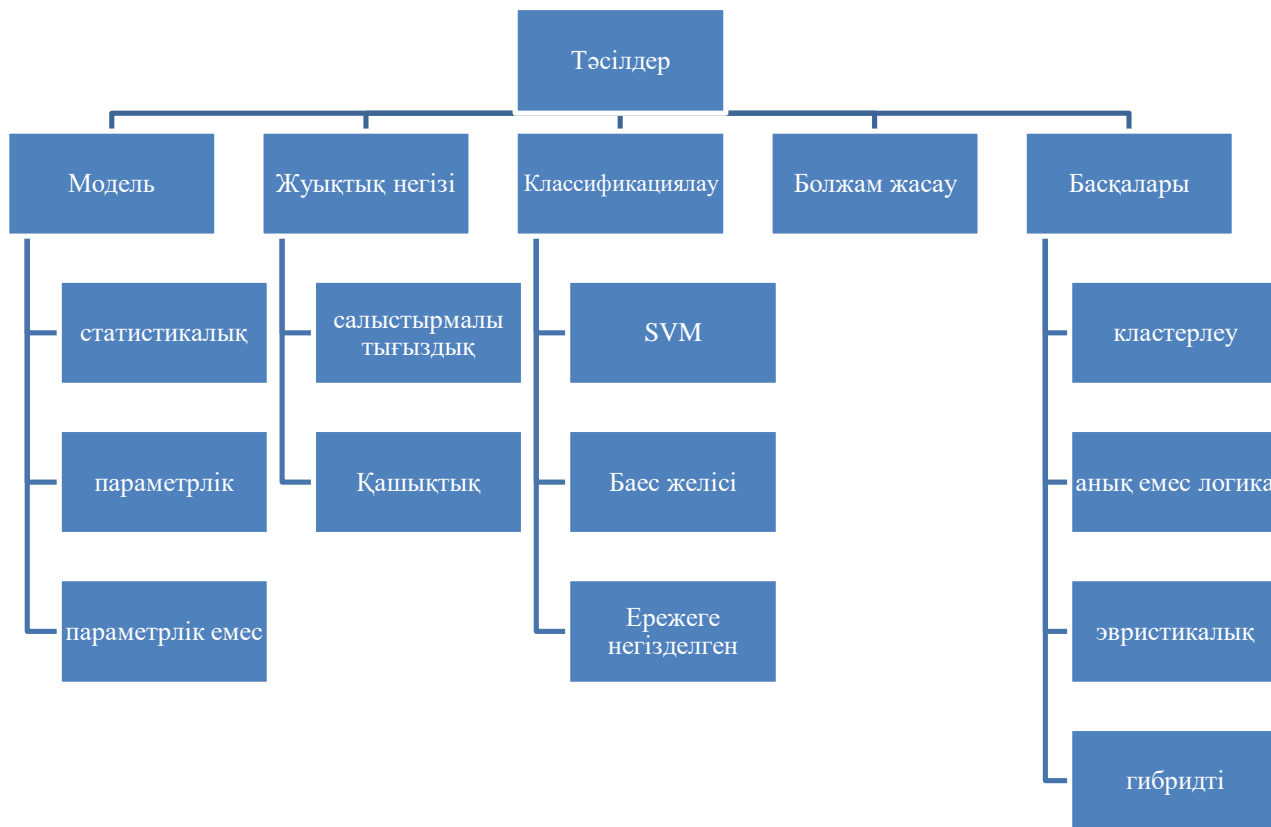
Үлкен деректердегі қайшылықтарды анықтау мен жоюға қолданылатын әдістер

Деректерді тазалау кезеңінде «дұрыс» деректерден сәйкес келетін ережелер немесе үлгілер қоры құрылады. Белгілі үлгі немесе ережеге сәйкес келмейтіндердің барлығы қайшылық немесе қате болып саналады. Деректердің үздіксіз өзгерісі жағдайында, деректердегі барлық мүмкін қателіктер мен қайшылықтар үшін үлгілер мен ережелер жасау мүмкін емес.

Барлық теңдестірілген қателерді жоя алмаймыз. Деректердің өзін өшіру арқылы ғана жоюға болатын қайшылықтар бар. Сондықтан, көп жағдайда деректерді мүлдем жоғалтқанша, оны «өңделмеген» түрде пайдалану туралы шешім қабылдаған жөн.

Қателерді жоюдан бас тарту. Үлкен деректерді талдаудағы маңызды фактор экономикалық тиімділік. Кейбір табылған қателерді жою процесі аса қиын. Егер қайшылықтарды жоюға кеткен шығындар алынған пайдадан көп болса, онда тазалаудан бас тарту шешімі қабылданады.

Үлкен көлемді деректердегі машиналық оқыту әдісінің қозғаушы күші оның интеллектуалды автоматтандыру қабілеттілігі болады. Сонымен қатар, бұл адамдар немесе дәстүрлі әдістер арқылы талдаумен салыстырғанда жасырын үлгілерді нарық тенденциясы мен клиент қалауымен белгілі бір уақыт аралығында анықтауды жеңілдетеді. Нақты бір мәселені шешіп, деректерді өңдеуде машиналық оқытуды пайдалану үшін әртүрлі әдістердің мықты және әлсіз жақтарын түсіну керек. 1-суретте қайшылықты анықтауға қолданылатын тәсілдер жіктелуі қарастырылды.



1-сурет. Әр түрлі тәсілдер негізінде қайшылықтарды анықтау әдістерінің жіктелуі

Деректерді тазалау бағдарламаларының өздері де қате тудыру көздері болуы мүмкін. Мысалы, деректерді тазалау бағдарламасы бір жазбаны түзете келе оны басқа жазбаның көшірмесі немесе қарама-қайшылығына айналдыруы мүмкін. Мұндай жағдай бағдарлама жеткіліксіз мәндердің орнына, белгілі бір ережеге сәйкес таңдалатын (мысалы, бағана бойынша орташа мән) мәнді қойған кезде туындауы мүмкін. Кейде деректерді тазалау процедурасы жағдайды одан бетер қиындатып жібереді. Егер OLTP-жүйесінің операторы кішігірім қателер туғызса, тазалау алгоритмі «жоқ» қатені түзету мақсатында біртегізде мыңдаған жазбаларды түрлендіруі мүмкін [8].

Семантикалық жақын ақпараттардағы қайшылықтарды шешу мүмкіндігі мен нұсқалары туралы шешімді Мамдани алгоритмін немесе нейронды желілерді пайдаланып, айқын емес қорытындылар әдісінің негізінде қабылдауға болады [3-8].

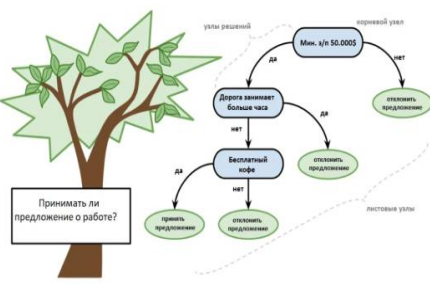
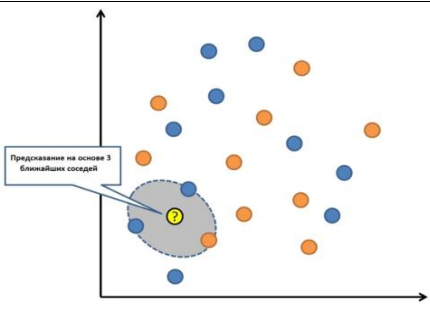
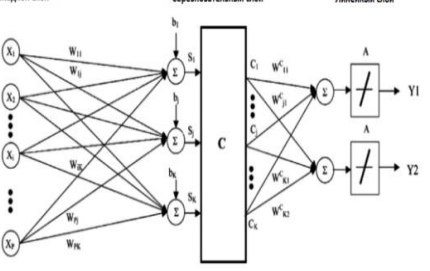
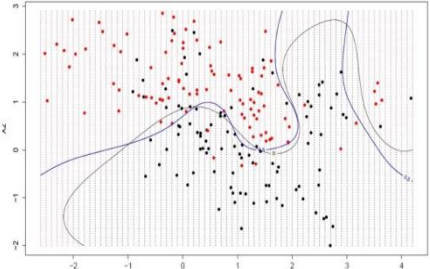
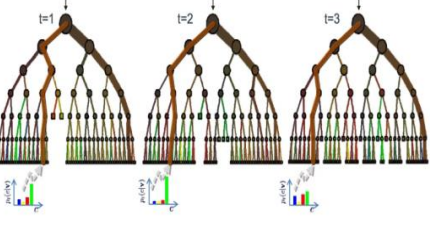
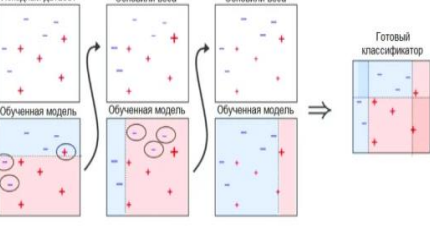
Үлкен деректер шеңберінде көбінесе машиналық оқыту әдісі мен желілік теория әдістері талқыланады. Талдау, бағалау және деректерді алу үшін Байес желісін қолданамыз. Деректер қайшылығы және осы мәселені шешуге арналған әдістерге басты назар аударылған. Машиналық оқытудың жалпы үш негізгі тәсілі бар: оқытушы көмегімен оқыту; оқытушы көмегінсіз оқыту; жартылай бақылаумен оқыту. .

Оқытушы көмегімен оқыту - бұл мақсатты айнымалы мәнді шешуге бағытталған машиналық оқытудың ең көп тараған түрі [9]. Оқытушы көмегімен оқыту көбінесе жіктеу және болжау жасау есептері үшін қолданады. 4-кестеде машиналық оқыту әдістерінің негізгі алгоритмдеріне қысқаша сипаттамалар берілген.

4-кесте. Машиналық оқыту әдістеріне сипаттама

№	Әдіс атауына түсінік	Графикалық түрде берілуі	Сипаттамасы
1	Аңғал Байес жіктелуі – қарапайым, бірақ өте тиімді алгоритм.		Модель екі ықтималдық типінен тұрады: 1. Әр класстың ықтималдылығы; 2. x тің әрбір мәніндегі әрбір класс үшін шартты ықтималдылық
2	Сызықтық регрессия – машиналық оқытудағы және статистикадағы аса танымал және түсінікті алгоритмдердің бірі		Сызықтық регрессия-кіріс x және шығыс y арасындағы байланысты нақты теңдеу түрінде көрсетеді. берілген
3	Логистикалық регрессия – статистикадан тұру машиналық оқытуға келген алгоритмдердің бірі. Оны бинарлық жіктеуге есептері үшін қолданған жақсы.		Кіріс айнымалылар үшін коэффициенттер мәндерін табу бұл жерде логистикалық регрессия сызықтыққа ұқсас боп келеді.
4	Егер класстар екіден артық болса, онда LDA (Linear discriminant analysis) алгоритмін қолданады. LDA алгоритмі әрбір классқа есептелген деректердің статистикалық қасиеттеріне ие.		Әрбір кіріс айнымалысы үшін: әр класс үшін орташа мәні, барлық класс бойынша дисперсияны есептеу қамтиды

• **Технические науки**

<p>5</p>	<p>Шешімдер бұтағы құрылымдық деректер мен алгоритмдер бойынша белгілі екілік бұтағы ретінде ұсынуға болады. Әрбір торап кіріс айнымалысы мен осы айнымалыны бөлу нүктесін білдіреді.</p>		<p>Жапырақтық тораптар - болжам үшін қолданылатын шығыс айнымалылары. Бұтақтар тез оқытылады және болжам жасайды.</p>
<p>6</p>	<p>К жақын көршілер әдісі өте қарапайым және өте тиімді алгоритм.</p>		<p>kNN деректерді сақтау үшін көп орын талап еткенімен болжамды тез жасайды. Сонымен қатар уақыт ағынымен болжамның нақты болуын қалдыру үшін оқытылатын деректерді жанарту отыру қажет.</p>
<p>7</p>	<p>Векторлық кванттау желісі LVQ (Learning vector quantization) векторлар коды жиынтығын ұсынады.</p>		<p>KNN алгоритмі сияқты оқытудан кейін векторлар болжам жасауға қолданады. Деректердің жаңа көшірмесі мен векторлар коды арасындағы аралықты есептеу арқылы алгоритм ең жақын көршіні табу.</p>
<p>8</p>	<p>Тірек векторлар әдісі машиналық оқыту-дағы көпшілікке мәлім және талқылауға түсетін алгоритмдерінің бірі. Гипержазықтық кіріс айнымалыларын кеңістікте бөлетін сызық.</p>		<p>0 мен 1 класстары бойынша ең тиімді қалыпта кеңістікте кіріс айнымалы-ларын бөлу тірек векторлар әдісінде гипержазықтық таңдалады.</p>
<p>9</p>	<p>Кездейсоқ орман - машиналық оқытудың ең танымал және тиімді алгоритмі. Оны бэггинг деп атауға да болады.</p>		<p>Кездейсоқ орман алгоритмінде оқыту мәліметтерінен алынған барлық үлгілер үшін шешім ағаштары жасалады. Егер шешімдер бұтағы жақсы нәтиже берсе осыны одан да жақсырақ ету үшін бэггингті қолданады.</p>
<p>10</p>	<p>Бустинг – бұл ансамбльдік алгоритмдер топтамасы, оның мәні бірнеше әлсіздерге негізделген күшті классификатор құру</p>		<p>Барлық ағаштар салынғаннан кейін жаңа мәліметтер үшін болжамдар жасалады және әр ағаштың тиімділігі оның жаттығу деректерінде қаншалықты дәл болғанына байланысты анықталады.</p>

5-кесте. Қайшылықты анықтаудағы қолданылатын машиналық оқыту әдістерінің мысалдары

Оқыту әдістері	Әдістер	Қолданбалы мәнмәтін
Басқарылатын	Марковтың жарысын моделі (Hidden Markov Model HMM) [10]	Басқарылатын статистикалық марковтық модель, онда модельденген жүйе Марковтік жасырын күй процесі болып саналады: қайшылықтарды анықтау үшін қолданылады [11, 12].
	Тірек векторлар әдісі (Support Vector Machine (SVM)) [13]	Деректер нүктелерінің кеңістіктегі көрінісі жекелеген санаттар олардың нақты бөлінуіне бөлінетін етіп көрсетілген: аномалияларды анықтау үшін SVM-дің арнайы класы, атап айтқанда бір SVMs (OCSVMs) кеңінен қолданылды [14].
	Гаусстық регрессия (GR) [15]	Регрессиялық және ықтималды жіктеуді шешуге арналған жалпы басқарылатын оқыту әдісі мәселелері: [16, 17] -де бейнеден қайшылықтарды анықтау үшін қолданылады.
	Нейрондық желілер (Convolutional Neural Networks CNN) [18]	Әдетте көрнекі бейнелерді талдауда қолданылатын терең нейрондық желілер класы: оның мағыналық деңгей функцияларын кірістен шығаруда қолдануға болатындығына байланысты, ол көптеген қосымшаларда танымал болды, соның ішінде қайшылықтарды анықтау [19, 20].
	Ұзақ мерзімді қысқа мерзімді есте сақтау (LSTM) сети [21]	Уақыттық қатарлар қосымшаларында қолданылатын қайталанатын рекуррентті нейрондық желілердің (RNN) ерекше түрі: [22, 20.] Ол қайшылықтарды анықтауға пайдаланылды.
	Жылдам аймаққа негізделген -CNN (Быстрый R-CNN) [24]	Кәдімгі CNN-ге сәйкес жіктелген нысанда тиімді жұмыс істейтін нейрондық терең нейрондық желілердің (DNN) жоғарырақ вариациясы: [25] -де ауытқуларды анықтау үшін қолданылады.
Бақылаусыз	Жасырын таратылған Дирихле (LDA) [26]	Құжаттардағы тақырыптардың негізгі бөлінуін алу үшін статистикалық талдауды қолдану тақырыптық моделі. Видеодағы визуалды сөздерді моделдеу үшін ауытқушылықты анықтау қолданылады [27].
	Ықтималды латентті семантикалық талдау (pLSA) [28]	Ықтималдық құрылымында бірлескен пайда болуы туралы ақпарат ұсынудың үлгісі: ауытқушылықтарды анықтау үшін қолданылады [29]
	Дирихленің иерархиялық процесі (HDP) [30]	Деректерді кластерлеу үшін LDA негізінде құрылған параметрлі емес Байес тәсілі: деректерді модельдеуде қайшылықты анықтауда қолданылады [31].
	Фишер әдісі [32]	Екі нысанның ұқсастығын өлшеу функциясы әр объект үшін өлшемдер жиынтығы мен статистикалық модельге негізделген: ол траекторияның көрінісін алу үшін қолданылады [33].
	Негізгі компоненттерді талдау (PCA) [34]	Бақылау жиынтығын түрлендірудің ортогональды түрлендірудің статистикалық процедурасы, мүмкін, корреляцияланған емес айнымалыларды сызықтық реттелмеген айнымалы шамалардың жиынтығына айналдырады: өлшемін азайту үшін қолданылады [35].
Гибридті	HDP + HMM	Гибридті модель: MIL көмегімен аномалияларды анықтау үшін [36] ішінде ішкі траекторияларды көрсету үшін қолданылады
	GAN-LSTM [35]	Гибридті модель: [20] пайдаланылған матчта жаттығуға қажет жалған жақтаулар екі бағытты Conv-LSTM көмегімен жасалады [35]
	CNN-LSTM	Гибридті модель: CNN-LSTM көмегімен болжауға негізделген аномалияны анықтау

Қорытынды. Үлкен деректерді өңдеу арқылы қажетті ақпаратты алу процесі бірнеше кезеңнен тұратындығы белгілі. Солардың бірі қайшылықты деректерді анықтап оларды жою әдістерінің алатын орны мол. Қазіргі уақыттағы замануи технологияларды қолдана отырып тиімділігін арттыру алға қойған мақсаттардың бірі. Мақалада үлкен деректерде кездесетін қайшылықты деректерді анықтап, оларды жою әрекеттеріне тоқталамыз.

ӘДЕБИЕТТЕР

- [1] <https://kk.wikipedia.org/wiki/%D2%9A%D0%B0%D0%B9%D1%88%D1%8B%D0%BB%D1%8B%D2%9B>
- [2] García S, Luengo J, Herrera F. Предварительная обработка данных в области интеллектуального анализа данных. Берлин: Спрингер; 2015. Посмотреть статью Google Scholar – //электронная книга
- [3] Han J, Kamber M, Pei J. Data Mining: концепции и методы, 3-е изд. Берлингтон: Morgan Kaufmann Publishers Inc; 2011. MATHGoogle Scholar
- [4] Quinlan JR. C4.5: Программы машинного обучения. Сан-Франциско, Калифорния: Morgan Kaufmann Publishers Inc.; 1993. Google Scholar
- [5] Agrawal R, Srikant R. Быстрые алгоритмы для правил ассоциации шахт. В: Материалы 20-й конференции по очень крупным базам данных (VLDB): 1994. с. 487-99. Google ученый
- [6] <https://www.researchgate.net/publication/261448807>
- [7] D. Zhang, On Temporal Properties of Knowledge Base Inconsistency. *Springer Transactions on Computational Science V*, LNCS 5540, 2009, pp.20-37.
- [8] A. Rodriguez, Inconsistency issues in spatial databases, in L. Bertossi et al (eds.) *Inconsistency Tolerance*, LNCS 3300, Springer-Verlag, 2004, pp.237-269.
- [9] <http://www.swsys.ru/index.php?page=article&id=3483>,
- [10] S. Biswas and R. V. Babu. Short local trajectory based moving anomaly detection. In ICVGIP, 2014.
- [11] T. Wang, M. Qiao, Y. Deng, Y. Zhou, H. Wang, Q. Lyu, and H. Snoussi. Abnormal event detection based on analysis of movement information of video sequence. *Optik-International Journal for Light and Electron Optics*, 152:50–60, 2018.
- [12] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.
- [13] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [14] C. E. Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.
- [15] K. Cheng, Y. Chen, and W. Fang. Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. *IEEE Transactions on Image Processing*, 24(12):5288–5301, Dec 2015.
- [16] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette. Real-time anomaly detection and localization in crowded scenes. In *CVPRW*, 2015.
- [17] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [18] X. Hu, S. Hu, Y. Huang, H. Zhang, and H. Wu. Video anomaly detection using deep incremental slow feature analysis network. *IET Computer Vision*, 10(4):258–265, 2016.
- [19] J. R. Medel and A. Savakis. Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390, 2016.
- [20] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [21] W. Luo, W. Liu, and S. Gao. Remembering history with convolutional lstm for anomaly detection. In *ICME*, 2017.
- [22] N. Srivastava, E. Mansimov, and R. Salakhudinov. Unsupervised learning of video representations using lstms. In *ICML*, 2015.
- [23] R. Hinami, T. Mei, and Shin'ichi S. Joint detection and recounting of abnormal events by learning deep generic knowledge. In *ICCV*, 2017.
- [24] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [25] H. Jeong, Y. Yoo, K. M. Yi, and J. Y. Choi. Two-stage online inference model for traffic pattern analysis and anomaly detection. *Machine vision and applications*, 25(6):1501–1517, 2014.
- [26] Hofmann. Probabilistic latent semantic analysis. In *UAI*, 1999.
- [27] R. Kaviani, P. Ahmadi, and I. Gholampour. Automatic accident detection using topic models. In *ICEE*, 2015.
- [28] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *NIPS*, 2005.
- [29] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis. Multiple hierarchical dirichlet processes for anomaly detection in traffic. *Computer Vision and Image Understanding*, 169:28–39, 2018.
- [30] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *ECCV*, 2010.

- [31] J. Wang, L. Xia, X. Hu, and Y. Xiao. Abnormal event detection with semi-supervised sparse topic model. *Neural Computing and Applications*, Mar 2018.
- [32] S. Ji, W. Xu, M. Yang, and K. Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):221–231, 2013.
- [33] L.-L. Wang, H. Y. T. Ngan, and N. H. C. Yung. Automatic incident classification for large-scale traffic data by adaptive boosting svm. *Information Sciences*, 467:59–73, 2018.
- [34] W. Yang, Y. Gao, and L. Cao. Trasmil: A local anomaly detection framework based on trajectory segmentation and multi-instance learning. *Computer Vision and Image Understanding*, 117(10):1273–1286, 2013.
- [35] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NIPS*, 2015.

Касымова Д.Т., Ахмедиярова А.Т., Шаяхметова А.С., Тұрдалыұлы М.

Анализ методов выявления и устранения противоречий в больших данных

Резюме. В статье анализируются конфликты, связанные с обработкой больших объемов данных и методы, используемые для их разрешения. Сложной задачей является выявление противоречивых данных между большими данными. Проанализированы подходы к выявлению и устранению противоречивых данных. Рассматривались использование методов машинного обучения при работе с противоречивыми данными.

Ключевые слова: большие данные, противоречие, временное противоречие, пространственное противоречие.

УДК: 621.372.543

¹A.A. Nauryzbayeva, ¹A.Zh. Laubayeva, ¹A.K. Zhunussova,
¹M. B. Bakytova, ²D. I. Ussipbekova

(¹Kazakh National Agrarian University, Almaty, Kazakhstan, ²Almaty University of Energy and Communication, E-mail: aigerimnauryz@mail.ru)

RESEARCH PROBLEMS IN THE DEVELOPMENT OF NANOTECHNOLOGY IN ENGINEERING

Abstract: Over the past few years, a short word with great potential - "nano" quickly entered the world consciousness. The article deals with the study of the problem in the development of nanotechnology in mechanical engineering. Nanotechnology is now in the early stages of development, as the main discoveries predicted in this area have not made yet. However, the ongoing research is already given practical results. The use of advanced scientific results in nanotechnology allows us to refer it to high technologies.

Key words: nanotechnology, science, technology, nanoelectronics, nanomaterials, some problems in the development of nanotechnology.

¹A.A. Наурызбаева, ¹A.Ж. Лаубаева, ¹A.Қ Жунусова, ¹М.Б. Бақытова,
²Д.И. Усипбекова

(¹Қазақ ұлттық аграрлық университеті, Алматы, Қазақстан Республикасы, ²Алматы Энергетика және байланыс университеті, E-mail: aigerimnauryz@mail.ru)

МАШИНА ЖАСАУДАҒЫ НАНОТЕХНОЛОГИЯНЫҢ ДАМУЫ МЕН МӘСЕЛЕСІН ЗЕРТТЕУ

Андатпа: Соңғы бірнеше жылда үлкен әлеуеті бар қысқа сөз - "нано" әлемдік санаға тез кірді. Мақалада машина жасау саласындағы нанотехнологияларды дамыту кедергілерін зерттеу қарастырылған. Нанотехнология қазір дамудың бастапқы сатысында тұр, себебі бұл салада болжанатын негізгі ашылулар әлі жасалған жоқ. Дегенмен, жүргізілген зерттеулер қажетті практикалық нәтижелер берді.

Нанотехнологияда озық ғылыми нәтижелерді пайдалану оны жоғары технологияларға жатқызуға мүмкіндік береді.

Кілтiк сөздер: нанотехнология, ғылым, техника, наноэлектроника, наноматериалдар, нанотехнологияларды дамытудың кейбір мәселелері.

Нанотехнология - бұл электроникада, оптикада, энергетикада, машина жасауда немесе биофизикада, биоинженерияда және медицинада болсын, Органикалық емес және органикалық

<i>Жетенбаев Н.Т., Балбаев Ф.Қ.</i>	
РОБОТОТЕХНИКА ЖӘНЕ ӨНЕРКӘСІПТЕГІ ЖАСАНДЫ БҮЛШЫҚ ЕТТЕР.....	446
<i>Нұрғалиев М.К., Саймбетов А.К., Омарали Б.М., Құттыбай Н.Б., Тұкымбеков Д.Х., Досымбетова Г.Б.</i>	
ӨРТҮРЛІ СОҢҒЫ ҚҰРЫЛҒЫЛАРЫ БАР LORA ТЕХНОЛОГИЯСЫ НЕГІЗІНДЕГІ СЫМСЫЗ ҚАБЫЛДАП-ТАРАТҚЫШ ҚҰРЫЛҒЫЛАР.....	454
<i>Байсбай Д.К., Боранкулова А., С. Мынбаева, А.Б. Умирбекова А.С., Сатыбалды А.М.</i>	
САРЫСУ НЕГІЗІНДЕ ИТМҰРЫН ШӘРБАТЫН ДАЙЫНДАУ.....	460
<i>Даулетбаева М.М., Джамилова С.М., Зубова О.А., Қожахан А.К.</i>	
УРБАНДАЛҒАН ТЕРРИТОРИЯДАҒЫ ШУДЫҢ ӘСЕРІ.....	464
<i>Ералиева Б.Ш., Тунгатарова А.Т., Боранкулова Г.С.</i>	
ЕНГІЗІЛГЕН ТАЛШЫҚТЫ ОПТИКАЛЫҚ ДАТЧИКТЕРДІҢ ИНТЕЛЛЕКТУАЛДЫ КОМПОЗИТТІК ҚҰРЫЛҒЫЛАРЫН ЗЕРТТЕУ.....	468
<i>Зәкенов С.Т., Нұршаханова Л.К. Тұрғияз М.Б.</i>	
ӘР ТҮРЛІ ӨНДІРУ ТӘСІЛДЕРІМЕН ЖАБДЫҚТАЛҒАН ҰҢҒЫМЕН ҚАБАТТАРДЫ БІР МЕЗГІЛДЕ-БӨЛЕК ПАЙДАЛАНУ ТЕХНОЛОГИЯСЫН ҚОЛДАНУ ТӘЖІРИБЕСІ.....	473
<i>Касимов А. О., Мұкаева А. Т.</i>	
ЖЕР СЕРІКТІК БАРЛАУ ЖҮЙЕЛЕРІНДЕ БӨГЕУІЛ ТҰРАҚТЫЛЫҒЫН ЗЕРТТЕУ.....	479
<i>Қасымова Д.Т., Ахмедиярова А.Т., Шаяхметова А.С., Тұрдалыұлы М.</i>	
ҮЛКЕН ДЕРЕКТЕРДЕ КЕЗДЕСЕТІН ҚАЙШЫЛЫҚТАРДЫ АНЫҚТАУ МЕН ЖОЮҒА ҚОЛДАНЫЛАТЫН ӘДІСТЕРГЕ ТАЛДАУ.....	486
<i>Наурызбаева А.А., Лаубаева А.Ж., Жунусова А.Қ., Бақытова М.Б., Усипбекова Д.И.</i>	
МАШИНА ЖАСАУДАҒЫ НАНОТЕХНОЛОГИЯНЫҢ ДАМУЫ МЕН МӘСЕЛЕСІН ЗЕРТТЕУ.....	494
<i>Заурбеков Т.Т., Досалиев К.С., Босак В.Н., Жантасов М.К., Алтыбаев Ж.М., Исмаилов Б.А., Жантасова Д.М.</i>	
ТАЛШЫҚТЫ-ЦЕМЕНТ ӨНДІРІСІН ОРЫНДАҒЫ АУАДАҒЫ ШАҢ КОНЦЕНТРАЦИЯСЫН АНЫҚТАУ ӘДІСТЕМЕСІ.....	497
<i>Широков И.А., Неверова Е.Г.</i>	
ЖЕКЕ МӘЛІМЕТТЕРДІ ҚОРҒАУ МАҚСАТЫНДА ИННОВАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРДЫҢ ҚОЛДАНЫЛУЫ.....	500
<i>Жуманов М.А., Щанова Б.Б.</i>	
ЖЭО ЖЫЛУЛЫҚ ЭНЕРГИЯ БЛОКТАРЫНДАҒЫ ЖЫЛУДЫ ТАЛДАУ ЖӘНЕ ЖҰМЫСЫН ЖЕТІЛДІРУ.....	504
<i>Гумиров Т.Ш., Тлемс А.Т., Тюканько В.Ю., Савинкин В.В., Дюрягина А.</i>	
СИЛИКОНДЫ БОЯУЛАР МЕН ЛАКТАРДАҒЫ КОМПОНЕНТТЕРДІҢ СУЛАНУЫН ЖАҚСARTУ АРҚЫЛЫ ЖАБЫННЫҢ ҚОРҒАНЫС РЕСУРСЫН АРТТЫРУ.....	510
<i>Джомартова Ш.А., Карымсакова Н.Т., Турсынбай А.Т., Жолмагамбетова Б.Р.</i>	
ХИМИЯЛЫҚ РЕАКТОРДЫ БАСҚАРУ ҮШІН АРАЛЫҚ ТАЛДАУДЫ ҚОЛДАНУ.....	516
<i>Каналы Т.С., Есенгараев Е.К., Телков Ш.А.</i>	
АЛТЫНҚҰРАМДЫ КЕНДЕРДІ ФЛОТАЦИАЛЫҚ БАЙЫТУ ҮШІН ОРГАНИКАЛЫҚ ӨНІМНЕН РЕАГЕНТТІ ҚОЛДАНУ.....	521
<i>Қуттыбаева А.Е., Кунтуган Е.Т.</i>	
СПУТНИКТІК ЖӘНЕ ОПТИКАЛЫҚ АРНАНЫҢ ӨТКІЗУ ҚАБІЛЕТТІЛІГІН ЗЕРТТЕУ.....	527
<i>Қуттыбаева А.Е., Толтаева Т.Р.</i>	
ЖЕР СЕРІКТІК БАРЛАУ ЖҮЙЕЛЕРІНДЕ БӨГЕУІЛ ТҰРАҚТЫЛЫҒЫН ЗЕРТТЕУ.....	530
<i>Отегенов М.Т., Бейсенов Б.С., Сарыбаев Е.Е.</i>	
ШОЙЫН ҚҰЮҒА БАЛАМА РЕТІНДЕ ФИБРОПОЛОМИРБЕТОНДЫ БҰЙЫМДАРДЫ ЖАСАУДЫҢ ЭКОНОМИКАЛЫҚ АСПЕКТІЛЕРІ.....	533
<i>Сабанова Б.К., Иванова И.В.</i>	
АЛТЫАЯҚТЫ ӨРМЕКШІ-РОБОТ ҮШІН ҚОЗҒАЛЫС АЛГОРИТМНІҢ БӨЛІКТЕРІН ЕСЕПТЕУ.....	535
<i>Молдабаева Г.Ж., Саукен Т.Ж.</i>	
ФЕРРОҚОРЫТПА ӨНДІРІСІНІҢ ШЛАКТАРЫН ӨНДЕУ МӘСЕЛЕСІ.....	541
<i>Аманжулова Ж.А., Избаирова А.С.</i>	
ЖҮК БЕКЕТІНДЕГІ МАНЕВРЛІК ЖҰМЫСТАРДЫ ТИІМДІ БАСҚАРУДЫҢ МАТЕМАТИКАЛЫҚ НЕГІЗДЕРІ.....	545
<i>Хидолда Е., Юрченко А. В., Жонкешова К. С.</i>	
MATLAB/SIMULINK ОРТАСЫНДА ФОТОЭЛЕКТРЛІК ПАНЕЛЬДІҢ ЭНЕРГЕТИКАЛЫҚ СИПАТТАМАЛАРЫН МОДЕЛЬДЕУ.....	551
<i>Умирбекова А. С., Боранкулова А.С., Досмырза Ж.С., Боранкулова Г.С.</i>	
АШЫТҚЫСЫЗ ӨЗДІГІНЕН АШЫТЫЛҒАН СҰЙЫҚ ҚАРА БИДАЙ АШЫМАЛЫН ҚОЛДАНЫП ҚАРА БИДАЙ НАНЫН ДАЙЫНДАУ.....	557