

**Құрамында улы элементтері бар жерді тазарту кезінде болжау және шешімдер қабылдау үшін ақпараттық жүйесің жобалау
НҰРЖАНОВ ШЫҢҒЫС АСҚАРУҰЛЫНЫҢ
6D060200-Информатика мамандығы бойынша философия докторы
(PhD) дәрежесін алуға арналған диссертациясының
АННОТАЦИЯСЫ**

Мәселенің өзектілігі. Компьютерлік (ақпараттық) технология табиғатта болып жатқан үрдістерді зерттеуге кең мүмкіндіктер ашты. Математикалық модельдеу әдістері мен есептеу техникасының көмегімен серпінді технологиялар негізінде адам мен табиғаттың өзара іс-қимылының жаһандық экологиялық мәселелерін шешу үшін жаңа әдістер, модельдер, алгоритмдер мен технологиялар жасалады. Қазіргі таңда Машиналық оқыту (МО) регрессия мен жіктеудің статистикалық әдістерінің алгоритмдерін қолдана отырып, ақпараттық ғылым, экология, ауыл шаруашылығы сияқты қазіргі заманғы деңгейдегі ғылымның әртүрлі салаларында қолданылады. Ауыл шаруашылығындағы кең ауқымды химияландыру ластануға ғана емес, сонымен қатар ауыл шаруашылығы дақылдарының өнімділігінің төмендеуіне әкелді, өсімдік шаруашылығының рентабельділігін арттыру өндірілетін өнімнің экологиясына байланысты. Осыған байланысты нақты уақыт режимінде дақылдардың өнімділігін бақылау үшін ауыл- шаруашылығын цифрландыру, экологиялық таза өнімді шығару аграрлық секторда жетекші орынды алады. Экологиялық қауіпсіздікті қамтамасыз ету үшін қоршаған ортаның жай-күйі туралы үлкен деректерді жинау есебінен экологиялық мониторингтің ақпараттық жүйелері әзірленуде, мониторинг жүйелерінің тиімділігі дәстүрлі мәліметтер базасының (МБ) басқару жүйесімен шешілетін деректерді үздіксіз жинау, өңдеу және сақтау үшін пайдаланылатын ақпараттық технологиялардың белгілі дәрежесімен айқындалады. Қазақстанда антропогендік бұзылған экожүйелер мен олардың қалдықтарының ластану деңгейін есепке алудың бірыңғай компьютерлік ақпараттық МБ жоқ. Бірыңғай МБ құрудың маңыздылығы ел Үкіметіне өнеркәсіптік өндіріс қалдықтарын жағу үшін жоғары технологиялық құрылыстар салу сияқты экологияға байланысты стратегиялық маңызды мәселелерді шешуге, ең бастысы техногендік ластанудың нақты деңгейін, қоршаған орта мен халықтың денсаулығына қауіпті бағалауға мүмкіндік береді.

Антропогендік белсенділіктен туындаған қоршаған ортаны ластаудың көлемінің ұлғаюына байланысты экожүйенің ластанған УЭ-н қалпына келтіру процесін математикалық модельдеуге қызығушылық жылдан жылға артып келеді. Ластану көзін тез бейтараптандыру алгоритмдері және топырақтың ластану үрдісінің сипаттамасын табуға, топырақ үрдістерінің бұзылуының салдарын, олардың өсімдіктер мен жануарлар әлеміне әсеріне оңтайлы шешім табуға, қалпына келтіру стратегиясын таңдауға мүмкіндік беретін математикалық модельдер жасалады. Математикалық модельдеуді қолданатын жүйелік талдау ластанған жерлерді, негізінен көмірсутектерді қалпына келтіру

үрдісін модельдеуге бағытталған. Сонымен қатар, топырақ қабатында ластанған аумақтарды улы элементтермен (УЭ) тазарту технологиясын жасауға негіз болатын басқа да негізгі ластаушыларды сипаттайтын және олардың "топырақ-өсімдік" жүйесіндегі көші-кон үрдісін сипаттайтын теория мен модель жоқ. Осыған байланысты, УЭ-мен ластанған аумақтарды қалпына келтіру технологиясы үшін ақпараттық жүйені әзірлеу, ластанған аумақтар туралы бірыңғай МБ және қалпына келтіруге қабілетті өсімдіктер туралы МБ құру қоршаған елді қорғау саласындағы маңызды экологиялық жобалар мен міндеттерді шешу, экологиялық қауіпті өңірлерде тұратын халықтың әлауқатын жақсарту үшін негіз болады.

Диссертациялық зерттеудің мақсаты – Қазақстан Республикасының жерлерін тазарту туралы болжау және шешім қабылдау үшін улы элементтермен ластанған топырақ туралы деректерді өңдеуге арналған зияткерлік ақпараттық жүйені әзірлеу.

Тапсырмалар

1. Улы элементтермен ластанған топырақтағы өсімдіктердің биоэнергетикалық өнімділігінің математикалық модельдеуді зерттеу.

2. Мәліметтер қорын құру: топырақты қалпына келтіруге қабілетті ағаш және шөптесін өсімдік түрлері туралы; аумақтың географиялық орналасуын ескере отырып, улы элементтермен ластанған жерлер туралы; ескірген пестицидтердің мөлшері мен олардың топырақтағы концентрациясы туралы.

3. Топырақты улы элементтерден тазарту процесін кезінде болжау және шешім қабылдаудың ақпараттық жүйесін әзірлеу.

4. Топырақтан улы элементтерді сіңіретін өсімдіктердің өнімділігін болжау үшін машиналық қыту әдістерін қолдану.

Зерттеу әдістері: Машиналық оқыту әдістері, регрессиялық теңдеулерді құру үшін өзін-өзі құрудың көп қатарлы эвристикалық әдісі, модельдеу әдістері, регрессиялық және дисперсиялық көп факторлы талдау әдістері. Ақпараттық жүйені құруға кешенді тәсілді әзірлеу кезінде ақпараттық жүйелерді жобалау теориясы, мәліметтер базасын жобалау әдістері, үрдіске бағытталған тәсіл қолданылды. Зерттеу нысаны ретінде улы элементтермен ластанған топырақтар мен өсімдіктер, климаттық жағдайлар туралы деректер (Алматы облысының мысалында) пайдаланылды. Зерттеу тақырыбы – `AlmatyWeatherDataSet.csv` климаттық деректері, 2015-2022 жж. улы элементтермен ластанған топырақтағы өсімдіктердің өнімділігі туралы деректер.

Зерттеудің негізгі нәтижелері:

1. Қоршаған орта жағдайына байланысты өсімдік биомассасын болжайтын «Өзін-өзі ұйымдастырудың көп қатарлы эвристикалық әдісі» бейімделді. Бұл процеске ең көп әсер ететін үш фактор: топырақ ылғалдылығының булануы, фотосинтездік белсенді сәулелену және жауын-шашын.

1.1 `Miscanmod` негізіндегі `Miscancalc` моделі ластанған және таза топырақ арасындағы айырмашылық коэффициентін есептеу арқылы климаттық деректерді ескере отырып, ластанған топырақтағы өсімдік биомассасын болжау үшін жақсартылды.

2. Екі ұйымдасқан деректер қоймасы құрылды: топырақты қалпына келтіруге ықпал ететін ағаш және шөптесін өсімдік түрлері бойынша; ескірген пестицидтердің мөлшері және олардың топырақтағы концентрациясы туралы.

3. Топырақты УЭ-ден тазарту процесі кезінде болжау және шешім қабылдау үшін ақпараттық жүйе әзірленді. Әрбір улы элементке жұмсалған тазарту кезеңінің ұзақтығы анықталды.

4. Интеллектуалды машиналық оқыту әдістері қолданылды және топырақтың ластану концентрациясын болжау үшін ең жақсысы анықталды. Зерттеу нәтижелері бойынша XGB Regressor төменгі метрикаға ие: $R2 = 0.998$; $MSE = 421.19$; $MAE = 15.03$; $MAPE = 0.065$.

4.1 Климат деректерінен өсімдік өнімділігін болжау үшін ансамбльді машиналық оқыту әдістері зерттелді. Талдау процесі Anaconda ортасында JupyterLab бағдарламалық құралы арқылы жүзеге асырылды.

4.2 Климаттық жағдайларға байланысты өнімділіктің регрессиялық үлгілері бағаланды. SHAP моделінің алгоритмі арқылы `datetime_1` (айлар) және `datetime_2` (күндер) белгілері жоғары дәрежелі әсер ететін өнімділікке жауап беретін 13 ақпараттық белгі таңдалды.

4.3 Әртүрлі элементтер үшін мискантус өсімдігінің көмегімен 0-20 см тереңдіктегі 1 га топырақ үшін топырақты тазалау кезеңінің ұзақтығы белгіленді. Ерекше улы элементтерге: қорғасынға – 12 жыл, мырышқа – 7 жыл.

Алынған нәтижелердің маңыздылығы мен жаңалығына негіздеме:

Келесі тұжырымдарды әзірлеу және алу *ғылыми жаңалық* болып табылады:

Климат деректерін ескере отырып, УЭ-мен ластанған топырақтағы өсімдік биомассасын болжау үшін Miscanmod негізіндегі жетілдірілген MiscanCalc моделі ұсынылады.

Болжамдық қасиеттері ақпараттық кіріс айнымалыларды автоматты түрде таңдауды және оңтайлы күрделіліктің регрессиялық моделінің құрылымын таңдауды қамтамасыз ететін регрессиялық модельдерден дәлдігі бойынша жоғары көп қатарлы өзін-өзі ұйымдастыру әдісі бейімделді.

Машиналық оқыту әдістеріне негізделген топырақты қалпына келтіру моделі жоғары өнімділікке және қайта оқуға төзімділікке ие XGBoost кітапханасының біріктірілген тәсілін қолдану арқылы кеңейтілді.

Нәтижелердің маңыздылығы: климаттық деректерді ескере отырып, УЭ ластанған топырақтағы өсімдік биомассасын болжауға арналған алгоритмдер жоғары математикалық дәлдікке ие. Ұсынылған көп ядролы өзін-өзі ұйымдастыру алгоритмі ақпараттық кіріс айнымалыларын автоматты түрде таңдауды және оңтайлы күрделіліктің регрессиялық моделінің құрылымын таңдауды қамтамасыз ете отырып, өзінің болжамды қасиеттері бойынша регрессиялық модельдерден асып түседі.

Нәтижелердің теориялық және практикалық маңыздылығы – табиғатты пайдалану мәселелеріндегі ақпараттық технология мен математикалық модельдің іргелі аспектілері. Қолданбалы құндылығы – рекультиви-

вациялық іс-шараларды қалыптастыру кезінде жер қорларын басқаруды жүзеге асыратын мемлекеттік органдарға ақпарат беру жеделдігінде; ақпараттық технологияларды енгізумен айналысатын тәжірибелі шаруашылықтардың, сондай-ақ агроэкологиялық мониторингті жүзеге асыратын ұйымдардың пайдалану мүмкіндігінде.

Ғылымды дамыту бағыттарына немесе мемлекеттік бағдарламаларға (жобаларға) сәйкестігі. Диссертациялық жұмыс ҚР ҰЭМ Ақпараттық және есептеу технологиялары институты негізіндегі бағдарлама және грант АР19678926 «Деректану әдістерін пайдалана отырып, топырақ пен атмосфералық ауаның ластануының экологиялық проблемаларын зерттеу және шешудің интеллектуалды жүйесін әзірлеу» (2023-2025 ж.) шеңберінде орындалды. Қорғауға шығарылған диссертациялық жұмыстың барлық нәтижелерін докторант автордың өзі орындады және жинады. Сонымен қатар, зерттеудің негізгі нәтижелерін, талдауларды, модель дерді, бағдарламаларды және ақпараттық жүйелерді докторант жасады.

Негізгі нәтижелердің ішінде: MiscanMod моделінің жаңа қосымшасы ретінде климаттық жағдайларға байланысты УЭ ластанған және ластанбаған топырақтағы өсімдік шығымдылығын болжауға; климаттық жағдайларға байланысты өнімділікке машиналық оқытудың регрессиясының 13 моделін бағалауға; УЭ ластанған аумақтары туралы базаны және әр түрлі зиянды заттармен ластанған топырақты тазартуға қабілетті қазақстандық флора өсімдіктерінің деректер базасын құруға; УЭ-ден топырақты тазарту үрдісінде болжау және шешім қабылдау үшін ақпараттық жүйені құруға арналған MiscanCalc қосымшасы. Өсімдіктің биомасса динамикасының қоршаған ортаның климаттық жағдайымен байланысын талдау үшін өзін-өзі ұйымдастырудың көп қатарлы алгоритмін әзірлеу физика-математика ғылымдарының докторы, профессор Т.Ж. Мазаковтың басшылығымен жүргізілді.

Диссертацияның негізгі ережелері мен зерттеу нәтижелері әл-Фараби атындағы Қазақ ұлттық университеті Ақпараттық технологиялар факультетінің Компьютерлік ғылым кафедрасының ғылыми семинарларында; ҒК ҰЭМ ҚР Ақпараттық және есептеу технологиясы институтының Ғылыми кеңесінде; "Информатика мен есептеу технологияларының заманауи мәселелері" ЖЗШ Халықаралық ғылыми конференциясында баяндамалар жасалды және талқыланды. 28-29 маусым 2016; II International Conference on Modern Problems of Computer Science and Computer Technology, MSHE RK, September 27-30, 2017; 15th Int. Phytotechnology conference. October 1-5, 2018. University of Novi Sad, Serbia; профессор У.А. Түкеевтің 75 жылдығына арналған ақпараттық технологиялар саласындағы Халықаралық ғылыми конференциясы, 2021 жылғы 8 қазан және басқа да халықаралық конференциялар.

Докторанттың әрбір жарияланымды дайындауға қосқан үлесінің сипаттамасы

Жарияланған мақалалар мен ғылыми еңбектерде диссертация тақырыбы бойынша зерттеу нәтижелері сипатталған. Ғылыми жұмыс барысында 17 ғылыми жұмыс жазылды, оның ішінде Қазақстан Республикасы Ғылым және

жоғары білім министрлігінің Ғылым және Білім саласындағы бақылау комитеті ұсынған журналдарда 4 мақала; халықаралық конференция материалдарында 8 жарияланымдар, Thomson Reuters және Scopus дерекқорына кіретін журналдарда 5 мақала.

Диссертацияның құрылымы мен көлемі. Диссертация белгілеу мен қысқартудан, кіріспеден, бес тараудан, қорытындыдан, пайдаланылған әдебиеттер тізімінен және қосымшадан тұрады. Диссертацияның толық көлемі 62 сурет пен 18 кесте және 3 қосымшасы бар 191 бетті құрайды. Әдебиеттер тізімінде 268 атау бар.

Кіріспеде диссертацияның өзектілігі негізделеді. Жұмыстың мақсаты, нысаны және зерттеу пәні тұжырымдалған. Ғылыми жаңалығы мен практикалық маңыздылығы анықталды. Зерттеу нәтижелері сипатталған.

Диссертацияның бірінші тарауында деректерді сақтау технологиялары, экология, ауыл шаруашылығы және экологиялық биотехнология саласындағы модельдеу әдістері негізінде қоршаған орта факторларының экологиялық мониторингінің ақпараттық жүйесі туралы әдеби шолу қарастырылады.

Екінші тарауда өсімдіктердің өнімділігін бағалау үшін УЭ ластанған климаттық жағдайларға байланысты (Алматы облысының мысалында) УЭ ластанған топырақта өсірілген өсімдіктердің өнімділігін бағалау үшін (мискантустың биоэнергетикалық, ауылшаруашылық түрлерінің мысалында) үш тәсіл қарастырылды: 1) ML-дің 13 регрессиялық моделі үшін алгоритмдерді пайдалану; 2) MiscanMod үлгісінің модификациясы негізінде MiscanCalc үлгісі үшін қосымшаны құру; 3) «Регрессиялық теңдеулерді құру үшін өзін-өзі ұйымдастырудың көп эвристикалық әдісі» математикалық моделін әзірлеу.

Үшінші тарауда УЭ ластанған аумақтар мен техногендік ландшафттарды қалпына келтіруге қабілетті өсімдіктер туралы мәліметтер базасында ақпараттық жүйені құруға кешенді тәсілді әзірлеу нәтижелері келтірілген. Функционалды талаптар, бағдарламаны құру кезеңдері және олардың құрылысы келтірілген. Қалдықтардың пайда болуы, оларды жою және сақтау және топырақты қалпына келтіру туралы есептілікті қалыптастыру үшін ақпараттық жүйе жасалды.

Төртінші тарауда өсімдіктердің вегетативті мүшелерінде ауыр металдардың жиналуын және олардың топырақ түріне және қоршаған орта жағдайларына байланысты топырақ-өсімдік жүйесінде миграциясын сипаттайтын математикалық модельді құру туралы мәліметтер келтірілген.

Бесінші тарауда УЭ өсімдіктерінің көмегімен 0-20 см тереңдіктегі 1 га топырақты тазарту үрдісін болжау үшін ақпараттық жүйені құру туралы мәліметтер келтірілген. Ақпараттық жүйенің тұжырымдамасы интеграцияланған тәсілге негізделген: өлшеу деректерін жинау, беру, жинақтау және өлшеу мәліметтерін өңдеу, мәліметтер базасының мәліметтері, өсімдіктердің өнімділік модельдері, топырақтың фитоуыттылық модельдері, өсімдіктің УЭ сіңіру модельдері және олардың топырақта болуы.

Қорытындыда диссертацияда алынған негізгі нәтижелер тұжырымдалған.