

FTAMP: 20.53.19

БИДАЙ ӨНІМДІЛІГІНЕ ӘСЕР ЕТУШІ КЛИМАТТЫҚ ФАКТОРЛАРДЫ МАШИНАЛЫҚ ОҚЫТУМЕН АНЫҚТАУДЫҢ МОДЕЛІ

Л.К.Тәжібай¹, Г.Е. Мырзабекова², Г.Ж.Стыбаев³, Г.К. Муратова⁴,
¹⁻⁴ С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті,
Жеңіс даңғылы, 62, Астана, Қазақстан
(E-mail: ¹ tazhibai_lazzat@mail.ru, ² g.murzabekova@kazatu.edu.kz, ³ g.stybaye@kazatu.edu.kz,
⁴ g.mugk@mail.ru)

Аннотация. Бидай өнімділігін болжау үшін маңызды климаттық факторларды анықтау ауыл шаруашылығындағы өсімдік шаруашылығына арналған бейімделудің тиімді стратегияларын жасауға мүмкіндік береді. Бұл бидай өндірісіндегі климаттық өзгерістердің кері әсерін азайтуға көмектеседі. Бұл зерттеудің мақсаты бидай өнімділігіне әсер ететін климаттық факторларды анықтау үшін сызықтық модельдер, шешім ағаштары және бустинг модельдері сияқты машиналық оқыту модельдерін әзірлеу болып табылады. Зерттеу Ақмола облысы Ақкөл ауданының 2000-2022 жылдардағы бидай өнімділігі мен метеодеректеріне, соның ішінде ауа температурасына, ылғалдылыққа, жауын-шашынға, желдің жылдамдығы мен бағытына, топырақ бетінің температурасына және ауа ылғалдылығының жетіспеушіліне негізделген. Бұл мақалада болжамдық модельдерді әзірлеу үшін машиналық оқытудың Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, SVR, Gaussian Process Regressor алты сызықтық моделі, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor шешім ағашының төрт моделі және LGBM регрессоры, Xgboost екі бустингтік модельдері қолданылды. Нәтижесінде бидай өнімділігіне едәуір әсер етуші белгілер, олардың арасында шілдедегі ауа ылғалдылығының жетіспеушілігі, маусымдағы максималды ылғалдылық және тамыздағы ауа температурасы анықталды. Ең жақсы нәтижелерді шешім ағашының модельдері көрсетті. Алынған мәліметтер интерпретацияланды және ауылшаруашылық құрылымдарында шаруашылық шешімдер қабылдау кезінде қолданыла алады.

Түйін сөздер: машиналық оқыту, болжамдық модель, белгілер маңыздылығы, зияткерлік талдау, шешімдер қабылдау.

DOI: <https://doi.org/10.32523/3007-0155/bulmathenu.2024/2.2>

2000 Mathematics Subject Classification: 62J05; 62J07; 68T01

1. Кіріспе

Ауыл шаруашылығындағы басқару шешімдерді талдауға, болжауға және тиімдендіруге мүмкіндік беретін заманауи цифрлық технологияларды қолданусыз ауыл шаруашылығы өндірісін тиімді басқаруды елестету мүмкін емес. Заманауи цифрлық технологияларды, оның ішінде зияткерлік талдауды қолдану ауыл шаруашылығы өндірісін тиімді басқаруға мүмкіндік ашады. Ауыл шаруашылығындағы деректерді жинақтау және талдау технологиялары мәселені шешудің механизмдерінің бірі бола алады және фермерлерге дәлірек шешім

қабылдауды қолдау құралдарын ұсына алады [1]. Деректерді зияткерлік талдау әдістері цифрлық агроөнеркәсіптік кешеннің негізі болып табылады.

Арнайы бағдарламалармен үлкен деректерді талдау ауыл шаруашылығының өнімділікті болжау, дәнді дақылдарды егу немесе жинау үшін қолайлы кезеңді анықтау, тыңайтқыштар мөлшерін есептеу және көптеген тағы басқа мәселелерін шешуге мүмкіндік береді [2]. Осы саладағы әртүрлі ғалымдардың зерттеулері, деректерді зияткерлік талдау сияқты технологияларды қолдану жұмыс тиімділігінің жоғарылауына, шығындарды азайтуға және кәсіпорын үшін тәуекелдерді азайтуға әкелетін шешім қабылдауға оң әсер ететінін растайды [3].

Машиналық оқыту моделі зерттеу мәселесі мен зерттеу сұрақтарына байланысты сипаттамалық немесе болжамды болуы мүмкін. Сипаттамалық модельдер жиналған деректерден білім алу және не болғанын түсіндіру үшін пайдаланылса, болжамды модельдер болашақты болжау үшін қолданылады [4]. Машиналық оқытуды зерттеу өнімнің болжамдық моделін құру кезінде әртүрлі мәселелерге тап болады. Қойылған міндетті шешу үшін дұрыс алгоритмдерді таңдау өте маңызды, сонымен қатар, алгоритмдер мен негізгі платформалар үлкен көлемдегі деректерді өңдей алуы керек [5].

Машиналық оқыту - бұл компьютерлік жүйелердің үлгілер мен логикалық тұжырымдарға сүйене отырып, нақты нұсқауларсыз тапсырмаларды орындау үшін қолданатын алгоритмдер мен статистикалық модельдерді әзірлеу туралы ғылым.

Машиналық оқытуды қолдану белгілі бір әрекеттер тізбегін орындауды білдіреді және алынған модельдің сәттілігі тек оқу алгоритмін таңдауға ғана емес, сонымен қатар, реттіліктің әр кезеңін дұрыс орындауға да байланысты [6].

Біздің алдыңғы [7] жұмысымызда машиналық оқыту маңыздылықтары, болжамды модельді машиналық оқытумен құру процесінің сызбалық көрінісі мен машиналық әдістерді қолдана отырып жасалған болжамдық модельдер туралы қысқаша жарияланды.

Зерттеу жұмысында машиналық оқытумен бидай өнімділігінің болжамдық моделін құру процесі бірнеше кезеңдермен жүзеге асырылады. Алдымен, қойылған міндеттер анықталып, содан соң, деректердің жинағы таңдалынып, синтезделеді. Болжамдық модель құруда деректерді тазалау және қалыпқа келтіру, модельді таңдау мен модельдеу және нәтижелерді талдау маңызды кезеңдер болып саналады.

Климаттық модельдер жиынтығының сценарийлерінде климаттың өзгеруі өнімділікке үлкен теріс әсерін тигізеді. Зерттеу жұмысында Crane-Droesch A. [8] жартылай параметрлік нейрондық желілер (SNN) деп аталатын терең нейрондық желілері бар параметрлік статистикалық модельдері әзірленген. Маңызды белгілер мен оларды нәтижемен байланыстыратын функционалдық формалар туралы алдын-ала білімді қолдана отырып, SNN типтік нейрондық желілермен салыстырғанда статистикалық тиімділікті едәуір арттырады. Климаттық модельдер жиынтығының сценарийлерінде модельдер климаттың өзгеруінің өнімділікке үлкен теріс әсерін көрсеткен.

Келесі [9] жұмыс авторлары фермалардан жиналған ауылшаруашылық деректерін, соның ішінде заттар интернетінің сенсорларының нақты уақыттағы деректерін қолдана отырып, модельдеу үшін әртүрлі машиналық оқыту алгоритмдерін фермада өсіруге ең қолайлы дақылдарды ұсыну үшін қолданды. Деректер жиынындағы температура, ылғалдылық, рН және жауын-шашын белгілерін пайдалана отырып, олар Байес желісі (Bayesian Networks) мен кездейсоқ орман әдістері арқылы ең жоғары дәлдікке қол жеткізді.

Өнімділікті арттыруға бағытталған [10] зерттеуде авторлар жіктеу және регрессия есептері үшін қолданылатын параметрлік емес әдіс болып табылатын KNN алгоритмінің барлық аспектілерін қарастырған. Жалпы алғанда, бақылау және тестілеу деректері толық деректердің сәйкесінше 80 және 20 пайызы ретінде таңдалған. Кіріс белгілері ретінде топырақтың қоректік құрамы бойынша азот (N), фосфор (P), калий (K), рН және температура деректері зерттелген болса, шығыс белгілері ретінде дақылдың бір түрі және өнімділік нәтижесі алынған.

Ж. Tussupov және т.б. авторлар [11] жұмыстарында спектрлік жарықтық коэффициенттерін пайдалана отырып, спектрлік деректерді талдау және спутниктік кескіндерде зиянкестер мен аурулардың болуын жіктеу үшін логистикалық регрессия, экстремалды градиентті күшейту (XGBoost) және үйірткілі нейрондық желі (CNN) сияқты машиналық оқыту әдістері қолданылған. Зерттеу нәтижелері ауыл шаруашылығы дақылдарының зиянкестері мен қоздырғыштарының формальды көріністерін талдау және тексеру нәтижесінде өнімділікті арттыруға, зиянкестер мен аурулармен күресу шығындарын азайтуға, өндірістік процестерді оңтайландыруға мүмкіндік береді.

М. Yessenova және т.б. авторлар [12] аэроғарыштық кескіндерді өңдеу арқылы вегетациялық кезеңді анықтау жөніндегі мәліметтер мен спектрлік жарықтық коэффициентті пайдалана отырып, ауыл шаруашылығы дақылдарының дамуының сандық сипаттамаларын анықтау әдістерін, сондай-ақ суреттерді текстуралық өңдеу моделін әзірледі. Зерттеулер ауылшаруашылығы өнімдерінің өсуіне әсер ететін жағымсыз факторларды жою үшін ұсынылған шараларды қабылдауға мүмкіндік береді.

П.К. Куценогий және т.б. [13] авторлар орманды далалы аймақта орындалған ұзақ далалық тәжірибелер материалдарына сүйене отырып, ауа-райы мен агротехникалық факторлардың бидай мен арпаның өнімділігіне әсерін машиналық оқытудың Random Forest Classifier алгоритмін қолданды. Бұл авторлар жұмысында accuracy көрсеткіші тестілік таңдамасында дұрыс болжамдар санының тестілік мысалдарының жалпы санына қатынасы ретінде анықтаған.

Нейрожелілік модельдердің артықшылықтарын қолдана отырып, М. Ю. Архипова [14] жүргізген зерттеуінде нейрондық желі моделі басқа нейрондық модельдермен (көп қабатты перцептрон және үйірткілі нейрондық желі) және регрессиялық модельдермен салыстырғанда, дәлірек болжамды бағалаулар алуына мүмкіндік берді. Ауылшаруашылық дақылдарының өнімділігін болжау модельдерінің дәлдігін жақсарту үшін компьютерлік көру технологиялары қолданылған.

Біздің зерттеуіміз бидай өндіруді тиімдендіру және негізделген шешімдер қабылдау арқылы шығындарды азайту үшін машиналық оқыту алгоритмдерін қолданудың маңыздылығын көрсетеді. Нәтижелер зерттелген аймақтың метеостанциясынан жинақталған метеодеректері бойынша климаттық жағдайлардың бидай өнімділігіне әсерін машиналық оқытудың сызықтық алгоритм, шешімдер ағашы, бустинг алгоритмдерін қолдана отырып, болжау арқылы алынды. Бұл агроқұрылымдарға климаттық факторларды ескере отырып, бидай өнімділігін тиімдендіру үшін дұрыс шешім қабылдауға мүмкіндік береді.

Модельдер бидай өнімділігі мен метеодеректердің тарихи деректері негізінде әзірленді. Солтүстік Қазақстан өңіріндегі Ақмола облысы, Ақкөл ауданы бойынша жаздық бидай өнімділігін болжау моделін әзірлеу міндетін орындау үшін бидай өнімділігіне әсер етуші климаттық факторлардың метеодеректері жинақталып, қажетті форматқа түрлендірілді. Жұмыста 2000-2022 жылдар аралығындағы күнделікті климаттық деректерін жаздық бидайдың жылдық өнімділігімен байланыстыру үшін машиналық оқыту қолданылды.

Жаздық бидай ұзақ жарық күнді талап ететін, өсудің қысқа мерзімді дақылына жатады [15]. Жер көлемі бойынша бұл әлемдегі ең көп өсірілетін дақыл және әлемдегі халыққа тағамдық калория мен ақуыздың бестен бір бөлігін қамтамасыз етеді [16]. Оның жылдық өндірісі орта есеппен 753 миллион тоннаны құрайды [17].

Бидай дақылының жаһандық азық-түлік қауіпсіздігі үшін маңыздылығын, сондай-ақ бидай өндірісіне климаттық факторлар әсерінің ерекшелігін ескере отырып, жұмыста бидайдың өнімділігін болжау және бидай өнімділігіне әсер ететін климаттық факторларды анықтау үшін модельдер әзірленді. Модельдеу процесі машиналық оқытуда қолданылатын Python бағдарламалау тілінің кітапханаларымен жүзеге асырылды. Болжамдық модельдерге салыстырмалы талдау жүргізіліп, бидай өнімділігіне әсер етуші климаттық факторлардың маңыздылығы айқындалды.

2. Әдістер

Бұл жұмыста бидайдың өнімділігі туралы мәліметтер Қазақстан Республикасының Ұлттық статистикалық бюросынан (<https://stat.gov.kz/>), ал метеодеректер «Қазгидромет» РМК-дан (https://meteo.kazhydromet.kz/database_meteo/) алынды.

Әзірленген болжамдық модельдер үшін қолданылған бидай өнімділігі мен метеодеректер бойынша мәліметтер қоры келесі сілтеме бойынша сақталған (<https://drive.google.com/drive/folders/1JckA0xvIbLgXxerZF0N3iv1Gufgx0Tn6?usp=sharing>).

Машиналық оқытуда шешуші рөл атқаратын және жиі еленбейтін үлкен нәрсе - бұл модельдер үшін маңызды белгілерді (әсер етуші факторларды) таңдау болып табылады (feature importance). Пайдасыз деректер машиналық оқытудың соңғы нәтижелерін бұзатын ығысуға әкеледі.

Маңызды белгілер – бұл модельдің нәтижелеріне және оның дәл болжау жасау қабілетіне айтарлықтай әсер ететін айнымалылар. Өсімдік шаруашылығындағы басқару шешімдерін тиімдендіруде, айталық, бидай өнімділігі мен қуаңшылықты болжауда, маңызды белгілер ретінде ауа температурасы, жауын-шашын сияқты белгілерді алуға болады.

Белгілердің маңыздылығы келесі себептерге байланысты өте пайдалы болады:

1) Деректерді түсіну: Модельді құру өз алдына маңызды болса, оған қолданылатын деректерді түсіну тағы бір маңызды кезең болып табылады. Белгілердің маңыздылығы белгілер мен мақсатты айнымалы арасындағы байланысты түсінуге мүмкіндік береді. Ол сондай-ақ, модель үшін қандай белгілер маңызды емес екенін түсінуге көмектеседі.

2) Модельді жақсарту: Модельді үйрету кезінде модельдің өлшемін азайту үшін белгілердің маңыздылығына негізделген бағалауды қолдануға болады. Жоғары көрсеткіштер әдетте сақталады, ал төменгі көрсеткіштер модель үшін маңызды емес болатындықтан жойылады. Бұл модельді жеңілдетіп қана қоймай, оның жұмысын тездетіп, нәтижесінде модельдің өнімділігі артады.

3) Модельдің интерпретациялануы: Белгілердің маңыздылығы модельді басқа мүдделі тараптарға түсіндіру және жеткізу үшін де пайдалы. Әрбір белгілер үшін көрсеткіштерді есептеу арқылы қандай сипаттамалар моделдің болжамды күшіне көбірек әсер ететінін анықтауға болады.

Машиналық оқыту моделі – бұл белгілі бір үлгілерді тануға үйретілген файл. Модельдер мәліметтер жиынтығы негізінде талдау және оқыту үшін қолдана алатын алгоритм бойынша оқытылады. Зерттеу жұмысында машиналық оқытудың мұғаліммен оқыту тәсілі қолданылды. Мұғаліммен оқытудың кез-келген алгоритмінің мақсаты – шығындар функциясын анықтау және оны азайту болып табылады.

Жаздық бидай ерекшелігіне қарай, модельдің деректер жиынына өнімділікке әсер етуші факторлар ретінде, ауа температурасы, жылдық жауын-шашын мөлшері, топырақ температурасы, ылғалдылық, ылғалдылық жетіспеушілігі, желдің бағыты мен жылдамдығы бойынша метеодеректері алынды. Деректер жиынтығы машиналық оқыту алгоритмдеріне сәйкес форматқа түрлендірілді.

Зерттеу жұмысындағы әзірленген модельдер Python бағдарламалау тілінің pandas, numpy, sklearn, seaborn, matplotlib, т.б. әртүрлі кітапханаларын қолданумен әзірленіп, арнайы интерфейс тік ортада нәтижелері шығарылды.

Төмендегі көрсетілген формулалар негізінде машиналық оқыту бойынша sklearn кітапханасындағы арнайы функциялар арқылы бидай өнімділігіне әсер етуші маңызды белгілер анықталды.

Сызықтық регрессия моделі келесі (1) бойынша есептеледі және оның мақсаты- болжанған \hat{y} мәндері мен нақты y мәндері арасындағы айырмашылықты минимизациялайтын $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ мәндерін табу болып табылады:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

мұндағы y – тәуелді айнымалы (мақсатты белгі), зерттелінді модельдегі (жылдық) бидай өнімділігі; x_1, x_2, \dots, x_n - тәуелсіз айнымалылар (белгілер), модельдегі бидай өнімділігіне л.н.

әсер етуші факторлар; β_0 - бос мүше (тұрақты), модельдегі барлық факторлардың нөлдік мәндеріндегі негізгі мәнін білдіреді және деректер кеңістігінде регрессия сызығының ығысуын қамтамасыз ететін регрессия теңдеуінің маңызды бөлігі ретінде қызмет етеді; $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ – әр белгінің мақсатты функцияға әсерін көрсететін регрессия коэффициенттері, әрбір фактор бір бірлікке өзгерген кезде модельдегі бидай өнімділігі қаншалықты өзгередінін көрсетеді; ε - модель қателігі (шу).

Сызықтық регрессияның жетілдірілген түрленуі ретінде қолданылған Ridge-регрессия моделі машиналық оқытуды азайту мақсатында $L2$ регуляризацияны кірістіреді және ол регрессия коэффициенттерінің мөлшеріне айыппұл қосады, бұл мультиколлинеарлық жағдайда көмектеседі және модельдің жалпылау қабілетін жақсартады:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^m (y_i - \hat{y})^2 + \alpha \sum_{j=1}^n \beta_j^2 \right) \quad (2)$$

мұндағы $\hat{\beta}$ - регрессияның бағаланған коэффициенттерінің векторы; y_i - тәуелді айнымалының нақты мәндері; \hat{y}_i - тәуелді айнымалының болжанған мәндері; α - регуляризация дәрежесін бақылайтын регуляризация параметрі (гиперпараметр); β_j - регрессия коэффициенттері; m - бақылаулар саны; n - белгілер саны.

Lasso-регрессия (Least Absolute Shrinkage and Selection Operator) әдісі Ridge-регрессия тәрізді регуляризациялауды қолдана отырып, модельдің жалпылау қабілетін жақсарту үшін қолданылды. Lasso-регрессиясының Ridge-регрессиядан айырмашылығы – бұл әдісте регрессияның кейбір коэффициенттерін нөлге айналдыратын $L1$ регуляризацияны қолданылады. Lasso-регрессиясы үшін:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^m (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^n |\beta_j| \right) \quad (3)$$

қолданылатын формула.

Lasso-регрессиясы мен Ridge-регрессиясының мақсаты - жоғалту функциясын мейлінше азайту болып табылады және ол екі кезеңмен орындалады:

1 $\sum_{i=1}^m (y_i - \hat{y})^2$ модель қателігін (қателер квадраттарының сомасын) есептеу;

2 $\sum_{j=1}^n |\beta_j|$ ($L1$ регуляризация Lasso-регрессиясы үшін) және $\alpha \sum_{j=1}^n \beta_j^2$ ($L2$ регуляризация Ridge-регрессиясы үшін) регрессия коэффициенттерінің өлшемі үшін айыппұл мөлшерін есептеу.

Айыппұл сызықтық регрессияның жоғалту функциясына қосылады және бұл модельді болжау дәлдігі мен коэффициенттердің шамасы арасындағы мәмілеге келуге мәжбүр етеді. Lasso регрессиясы кейбір коэффициенттерді нөлге айналдыруы мүмкін, бұл белгілерді таңдау үшін пайдалы болады. Осылайша, Lasso регрессиясы қайта оқытуды болдырмауға көмектеседі және белгілердің санын азайта алады. Lasso регрессиясының аталған ерекшеліктері модельді түсіндіруді жеңілдетеді.

Elastic Net регрессиясы Lasso-регрессиясы мен Ridge-регрессиясының артықшылықтарын қолдануға мүмкіндік береді. Бұл регрессия әдісі коэффициенттердің мөлшерін бақылап қана қоймайды, сонымен қатар белгілерді таңдайды. Elastic Net регрессиясы келесі формуламен есептеледі:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^m (y_i - \hat{y})^2 + \alpha_1 \sum_{j=1}^n |\beta_j| + \alpha_2 \sum_{j=1}^n \beta_j^2 \right) \quad (4)$$

мұндағы α_1 және α_2 - шамалары сәйкесінше $L1$ және $L2$ регуляризациялау дәрежесін бақылайтын регуляризация параметрлері.

Elastic Net регрессиясы λ параметрі көмегімен $L1$ және $L2$ регуляризациялары арасындағы қатынасты икемді түрде реттеуге мүмкіндік береді:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^m (y_i - \hat{y})^2 + \lambda (\alpha \sum_{j=1}^n |\beta_j| + (1 + \alpha) \sum_{j=1}^n \beta_j^2) \right),$$

мұндағы

λ – регуляризацияның жалпы коэффициенті;

α – $L1$ және $L2$ регуляризациялары арасындағы теңгерімді бақылайтын параметр.

Жұмыстағы тірек векторлар әдісі (Support Vector Regression, SVR) ε -нан артық емес, сондай-ақ, максималды тегіс болып табылатын, шынайы мәндерден ауытқитын функцияны табу мақсатымен қолданылды. SVR әдісінің формуласы келесідей сипатталады:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*)$$

және

$$\begin{cases} y_i - (\mathbf{w} \cdot x_i + b) \leq \varepsilon + \xi_i \\ (\mathbf{w} \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, m. \end{cases}$$

шектеулері орындалады, мұндағы \mathbf{w} – әр-бір фактор салмақтарының (маңыздылықтарының) векторы; b – бос мүше; ξ_i, ξ_i^* – ε – түтік шекарасынан шығатын i -ші нүкте үшін қателерді ұсынатын өлшемдер; C – регуляризациялау коэффициенті; x_i – i -ші бақылау белгілерінің векторы; y_i – i -ші бақылаудың мақсатты функциясы; ε – ішінде қателерге айыппұл салынбайтын ε -түтік ені.

Сызықтық регрессия модельдері нөлдік мәндерді жақсы өңдейді, ал Lasso және Ridge регрессиясы модельдердің жалпылау қабілеттілігін белгілердің үлкен санында жақсарта алады.

Осы жұмыста үш түрлі модельдер типтері қарастырылды: сызықтық, шешімдер ағашы, және бустинг.

3. Нәтижелер мен талқылау

Машиналық оқыту модельдерінде бидай өнімділігіне әсер етуші әрбір факторлар (орташа температура, максималды температура, жел жылдамдығының орташа мәні, топырақ беті температурасы және т.б.) бойынша Ақмола облысы Ақкөл ауданының 2000-2022 жылдар аралығындағы күнделікті әрбір үш сағат сайын алынған деректерінің әр күн бойынша орташа, максималды, минималды мәндері қолданылды. Оқытылған машиналық модельдерге салыстыру жұмыстары жүргізіле отырып, талданды. Зерттеу жұмысындағы өңделген деректер алдын-ала екі бөлікке (оқыту және тестілеу), атап айтқанда, оқытуға арналған деректердің 70%-ына және тестілеуге арналған деректердің 30%-ына бөлінген.

Машиналық оқыту үшін деректерді 70% және тестілеу үшін 30%-ға бөлу модельдің өнімділігін сенімді бағалауға көмектесетін стандартты тәжірибе болып табылады. Оқыту таңдамасы (training set) модельді оқыту үшін пайдаланылса, ал тестілік таңдамасы (test set) модель өнімділігін бағалау үшін қолданылады. Тестілеу таңдамасы модельдің машиналық оқыту кезінде жаңа деректердегі өнімділігін бағалау үшін қолданылады. Модельдің өнімділігін бағалау үшін қолданылатын MAPE (Mean Absolute Percentage Error) орташа абсолютті пайыздық қатесі әрқашан тестілік таңдамасында есептеледі.

Осы жұмыста болжамдық модельдердің қаншалықты дәл екенін бағалау үшін орташа абсолютті пайыздық қате (MAPE) метрикасы қолданылды. Бұл қате түрі қатарлардың мәнімен және іргелес қатарлардағы қателермен салыстырғанда қаншалықты үлкен екенін бағалайды. Ол өте қарапайым интерпретацияланатын коэффициент және төмендетілген болжамдарға қарағанда шамадан тыс болжау жағдайларына өте қатты қарайды. Оны үлеспен немесе пайызбен өлшеуге болады.

Орташа абсолютті пайыздық қате (MAPE) машиналық оқытуда әдетте модельдердің болжамды дәлдігін өлшеу үшін қолданылады. Ол келесідей есептеледі:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \quad (5)$$

мұндағы Y_t - t уақыт сәтіндегі нақты мәндер; \hat{Y}_t - t уақыт сәтіндегі болжанған мәндер; n - бақылаулар саны.

MAPE - уақыт қатары мен регрессия есептеріндегі болжау дәлдігін бағалау үшін ең көп қолданылатын көрсеткіштердің бірі. Бидай өнімділігін болжау моделі Linear Regression үшін орташа абсолютті пайыздық қатені (MAPE) есептеу мысалы төменде келтірілген. Болжамдық модельдер жоғарыда айтып кеткендей, 2000-2022 жылдар аралығындағы 23 жылғы деректер бойынша әзірленгендіктен, n бақылаулар саны – 23-ке тең. Деректер базасындағы Ақкөл ауданы бойынша көрсетілген уақыт кезеңіндегі әр жыл бойынша бидай өнімділігінің нақты мәндері - Y_t мәндері болса, Python бағдарламалау тілінің sclearn кітапханасындағы модельдерге сәйкес арнайы функцияларды қолдана отырып, машиналық оқыту барысында бидай өнімділігінің болжанған мәндері \hat{Y}_t алынды.

Нақты және болжамды мәндер арасындағы орташа абсолютті пайыздық қатені Linear Regression моделі бойынша есептеу үшін деректер 1-кестеде берілген.

КЕСТЕ 1 – Linear Regression моделі үшін нақты және болжамды мәндері

Жылдар (t уақыт)	Y_t	\hat{Y}_t	Жылдар (t уақыт)	Y_t	\hat{Y}_t	Жылдар (t уақыт)	Y_t	\hat{Y}_t
2000	8,9	10,1	2008	4,0	4,5	2016	10,3	11,6
2001	11,7	13,2	2009	12,7	14,4	2017	9,3	10,5
2002	8,3	9,4	2010	5,8	6,6	2018	10,4	11,8
2003	8,0	9,0	2011	12,6	14,3	2019	9,8	11,1
2004	4,9	5,5	2012	5,8	6,6	2020	12,0	13,6
2005	7,0	7,9	2013	8,8	10,0	2021	9,2	10,4
2006	9,5	10,7	2014	8,3	9,4	2022	10,8	12,2
2007	9,0	10,2	2015	10,5	11,9			

Алдымен әр жыл үшін абсолютті пайыздық қателік

$$\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$$

түрінде есептелінеді, есептеудің келесі қадамында (1) арқылы барлық абсолютті пайыздық қателіктердің орташа мәнін анықтау арқылы Linear Regression моделі үшін орташа абсолютті пайыздық қателік

$$MAPE = \frac{1}{23} (0,1348 + 0,1282 + \dots + 0,1296) = \frac{1}{23} \cdot 3,0032 = 0,131$$

санына тең болады.

MAPE орташа абсолютті пайыздық қате болғандықтан, есептелген нәтижені 100%-ға көбейтіп, 13,1% пайыздық нәтижесін аламыз (2-кесте).

Зерттеу жұмысында қолданылған сызықтық, шешімдер ағашы, бустинг алгоритмдері бойынша әзірленген бірнеше модельдердің тиімділігін бағалауға мүмкіндік беретін болжамды дәлдік нәтижелері 2-кестеде келтірілген.

MAPE пайызбен көрсетілгендіктен, ол әртүрлі деректер жиынтығындағы немесе түрлі болжау объектілеріндегі модельдердің дәлдігін салыстыруды жеңілдетеді. Мысалы, MAPE мәні ең жоғарғы көрсеткішті берген Gradient Boosting Regressor моделі 18,1% құрайды, бұл болжамды мән мен нақты мән арасындағы орташа айырмашылық 18,1% құрайды дегенді білдіреді. MAPE мәні неғұрлым төмен болса, модель соғұрлым мәндерді жақсырақ болжай алады. Мысалы, Extra Trees Regressor моделі (12,2%) Xgboost моделіне (16,3%) қарағанда дәлірек.

Ақмола облысының Ақкөл ауданы бойынша нақты бидай өнімділігі зерттелетін 23 жылдық уақыт кезеңіндегі 4-14,7 ц/га аралығында өзгереді. Агрономиялық саланың ғалымдарының пікірлері бойынша бидай өнімділігін анықтаудағы болжамдық модельдердің қателігі 10-15% аралығында рұқсат етіледі.

Төмендегі 2-кестеден тек үш модель жақсы нәтиже көрсеткендігі байқалады: Extra Trees Regressor (12,2%), Linear Regression (13,1%) және Ridge Regression (13,6%).

Кесте 2 – Қолданылған алгоритмдердің болжамдық нақтылығының өлшемі ретіндегі орташа абсолютті пайыздық қате (MAPE) көрсеткіштері

Р.с.	Алгоритм	Модель атауы	Орташа абсолютті пайыздық қате (MAPE, %)
1	Сызықтық алгоритм	LinearRegression	13.1
2		RidgeRegression	13.6
3		LassoRegression	16.6
4		ElasticNetRegression	16.7
5		SVR	15.9
6		GaussianProcessRegressor	15.2
7	Шешімдер ағашы алгоритмі	DecisionTreeRegressor	16
8		RandomForestRegressor	15.4
9	Бустинг алгоритмі	GradientBoostingRegressor	18.1
10		ExtraTreesRegressor	12.2
11		LGBMRegressor	17.1
12		Xgboost	16.3

Жоғарыда келтірілген орташа абсолютті пайыздық қате нәтижелерін ескере отырып, аталған бірнеше алгоритмдердің ішінде салыстырмалы түрде ең тиімділері – сызықтық алгоритмдерден Linear Regression, Ridge Regression модельдері болса, ал бустинг алгоритмдерінің ішінен Extra Trees Regressor моделі тиімді және ұтымды енгізу үшін заманауи алгоритм болып табылады.

Әзірленген модельдерде қолданылған бидай өнімділігіне әсер етуші климаттық факторлар орташа температура (Temp_average), максималды температура (Temp_max), минималды температура (Temp_min), максималды желдің жылдамдығы (WindSpeed_max), жел жылдамдығының орташа мәні (WindSpeed_average), жел бағытының орташа мәні (WindDir_average), минималды ылғалдылық (Hum_min), максималды ылғалдылық (Hum_max), ылғалдылықтың орташа мәні (Hum_average), топырақ беті температурасының максималды көрсеткіші (SoilTemp_max), топырақ беті температурасының минималды көрсеткіші (SoilTemp_min), топырақ беті температурасының орташа көрсеткіші (SoilTemp_average), ылғалдылық жетіспеушілігінің орташа мәні (DefHum_average), жауын-шашын сомасы (Precipitation_sum), жылдық жауын-шашын сомасы (Annual_prec).

Жүргізілген зерттеу жұмысындағы өнімділікті болжау барысында маңызды деректерді талдау үшін келесідей: Linear Regression; Ridge Regression; Lasso Regression; Elastic Net Regression; SVR; Gaussian Process Regressor; Decision Tree Regressor; Random Forest Regressor; Gradient Boosting Regressor; Extra Trees Regressor; LGBM Regressor; Xgboost машиналық оқыту модельдері қолданылды.

Бидай өнімділігін болжаудағы климаттық факторлардың маңыздылығын анықтау үшін сызықтық, шешімдер ағашы және бустинг әдістері арқылы маңызды белгілердің бета-коэффициенттері мен белгілер маңыздылығының көрсеткіштері машиналық оқытудың арнайы функцияларымен анықталды.

Сызықтық регрессия модельдері үшін әр белгінің тәуелді айнымалыға (бидай өнімділігі) әсер ету күші мен бағытын көрсететін бета-коэффициенттері (β) есептелді. Белгілердің тәуелді айнымалыға әсер ету бағыты бета-коэффициенттің оң немесе теріс таңбасымен анықталады. Оң бета-коэффициенттер белгінің өнімділікке оң әсерін көрсетсе, ал теріс коэффициенттер л.н.

сәйкесінше теріс әсерді көрсетеді. Шешім ағашы және бустинг модельдері үшін әрбір белгінің модельді жақсартуға қосқан үлесін көрсететін белгілердің маңызды көрсеткіштері анықталды. Нәтижелер 3-кестеде келтірілген.

Бета-коэффициенттері әр тәуелсіз айнымалының (факторлар) тәуелді айнымалыға (бидай өнімділігі) қаншалықты және қай бағытта әсер ететіндігі туралы түсінік береді. Бұл сызықтық регрессия модельдерін интерпретациялаудың және модельдегі қандай факторлар ең маңызды екенін түсінудің маңызды аспектісі болып табылады.

Сызықтық регрессия моделі бойынша жылдық жауын-шашынның 0,035 оң коэффициенті жылдық жауын-шашынның бір бірлікке артуы (1 мм) бидай өнімділігінің 0,035 бірлікке (ц/га) артуына алып келетіндігін білдірсе, Elastic Net моделі бойынша тамыздағы орташа температураның -0,304 теріс коэффициентінде тамыздағы орташа температураның бір бірлікке артуы (1°C) бидай өнімділігінің 0,304 бірлікке (ц/га) кемуіне байланысты екендігін көрсетеді.

Decision Tree Regressor моделі бойынша ауа ылғалдылығы тапшылығының 0,774 көрсеткіші белгінің модельге, яғни модельді жақсартуға 77% үлес қосатындығын білдіреді. Осы көрсеткіш арқылы әрбір белгінің болжам үшін қаншалықты маңызды екенін бағалауға болады. Бұл алгоритм модельдері үшін белгілердің тәуелді айнымалыға әсер ету бағыты деген түсінік жоқ, сол себепті белгілер маңыздылығының көрсеткіштері тек оң мәндерді береді.

Төменде машиналық оқыту барысында анықталған маңызды белгілердің сызықтық алгоритм модельдері үшін бета-коэффициенттері мен шешім ағашы және бустинг модельдері үшін белгілер маңыздылығының көрсеткіштері берілген. Машиналық оқыту барысында әр модель бойынша он маңызды белгілер анықталды. Төменде әр модель бойынша ең жақсы алғашқы бес бета-көрсеткіштері мен белгілер маңыздылығының көрсеткіштері берілген.

Сызықтық регрессия моделі бойынша бидай өнімділігіне (салмағының кему ретімен) әсер ететін ең маңызды белгілер ретінде келесілерді анықтады: жылдық жауын-шашын (0,035), қарашадағы (0,026) және маусымдағы (0,024) жауын-шашын мөлшері, ақпандағы ең төменгі ылғалдылық (-0,015) және жел бағыты (0,015).

Ridge-регрессиялық моделі (2) арқылы бидай өнімділігіне әсер ететін ең маңызды белгілер ретінде келесілерді анықтады: жылдық жауын-шашын (0,038), маусымдағы (0,025) және қараша (0,024) жауын-шашын, ақпандағы ең төменгі ылғалдылық (-0,018) және қазан айында жел бағыты (-0,015).

Lasso регрессия моделі бидай өнімділігіне әсер ететін ең маңызды белгілер ретінде келесілерді анықтады: қарашадағы жауын-шашын (0,084), жылдық жауын-шашын (0,035), ақпандағы минималды ылғалдылық (-0,021), қазандағы жауын-шашын (-0,021) және қыркүйектегі жел бағыты (-0,018).

Elastic Net моделі бидай өнімділігіне әсер ететін ең маңызды белгілер ретінде келесі нәтижелерді берді: маусымдағы максималды ылғалдылық (0,572), тамыздағы минималды ылғалдылық (0,397), тамыздағы орташа температура (-0,304), қарашадағы жауын-шашын (0,263) және ақпандағы минималды ылғалдылық (-0,243).

Шешім ағаштарына негізделген барлық модельдер шілдедегі ылғалдың тапшылығын маңызды белгі ретінде көрсетеді: Decision Tree Regressor (0,774), Random Forest Regressor (0,071), Gradient Boosting Regressor (0,690), Extra Trees Regressor (0,690).

LGBM Regressor және Xgboost бустингтік модельдердің ішінде ең жақсы мәнді Xgboost градиентті бустингтік моделі көрсетті: тамыздағы температура (0,422), шілдедегі ылғал тапшылығы (0,407), ақпандағы (0,161), қарашадағы (0,004) және маусымдағы ауа температурасы (0,003).

Сызықтық регрессияға арналған β_n коэффициенттері әр белгінің (климаттық факторлар) тәуелді айнымалыға (бидай өнімділігі) әсер ету күші мен бағытын бағалау үшін қолданылады. Бірнеше белгілері бар сызықтық регрессияда β_n коэффициенттерін есептеу бірнеше қадаммен жүзеге асырылады.

Алынған нәтижелердің ішінен, бидай өнімділігіне ақпан айындағы жел бағыты бойынша бета-коэффициентінің 0,015 мәнін көрсеткен (3-кесте) сызықтық регрессия моделі негізінде β_1 коэффициентін есептеу мысалын қарастырайық.

Деректер жиыны бойынша Y мәндері 2000-2022 жылдар аралығындағы Ақкөл ауданы бойынша бидай өнімділігінің (ц/га) көрсеткіші. X мәндері - көрсетілген жылдар аралығындағы жел бағытының (град.) ақпан айы бойынша орташа мәндері.

$Y = [8,9; 11,7; 8,3; 8,0; 4,9; 7,0; 9,5; 9,0; 4,0; 12,7; 5,8; 12,6; 5,8; 8,8; 8,3; 10,5; 10,3; 9,3; 10,4; 9,8; 12,0; 9,2; 10,8];$

$X = [234,082; 186,192; 174,259; 184,000; 162,009; 161,058; 129,348; 179,438; 194,000; 105,661; 125,571; 149,826; 63,487; 195,679; 149,063; 154,094; 161,272; 185,629; 130,795; 160,487; 180,909; 177,536; 151,750];$

Бақылаулар саны $n = 23$ (2000-2022 жылдар аралығы, жалпы 23 жыл).

1. X және Y айнымалылары үшін μ_X және μ_Y орта мәндерін табу:

$$\mu_X = \frac{1}{23} \sum_{i=1}^{23} X_i = 160,702;$$

$$\mu_Y = \frac{1}{23} \sum_{i=1}^{23} Y_i = 9,026.$$

2. σ_X және σ_Y стандартты ауытқуларды есептеу:

$$\sigma_X = \sqrt{\frac{1}{23} \sum_{i=1}^{23} (X_i - \mu_X)^2} = \sqrt{\frac{26728,898}{23}} = 34,090;$$

$$\sigma_Y = \sqrt{\frac{1}{23} \sum_{i=1}^{23} (Y_i - \mu_Y)^2} = \sqrt{\frac{121,404}{23}} = 2,297.$$

3. X және Y мәндерін стандарттау әдісімен нормалау (Python бағдарламалау тілінің StandardScaler функциясы арқылы келесі формула бойынша есептеледі):

$$X_{standardized,i} = \frac{X - \mu_X}{\sigma_X};$$

$X_{standardized,i} = [2,1525; 0,7477; 0,3977; 0,6834; 0,0383; 0,0104; -0,92; 0,5496; 0,9768; -1,615; -1,031; -0,319; -2,852; 1,026; -0,341; -0,194; 0,0167; 0,7312; -0,877; -0,006; 0,5928; 0,4938; -0,263];$

$$Y_{standardized,i} = \frac{Y - \mu_Y}{\sigma_Y};$$

$Y_{standardized,i} = [-0,055; 1,1638; -0,316; -0,447; -1,796; -0,882; 0,2063; -0,011; -2,188; 1,5991; -1,404; 1,5556; -1,404; -0,098; -0,316; 0,6415; 0,5545; 0,1192; 0,598; 0,3369; 1,2944; 0,0757; 0,7721].$

4. β_1 коэффициентін есептеу (жел бағыты белгісі бойынша β_1 коэффициенті):

$$\beta_1 = \frac{\sum_{i=1}^{23} (X_{standardized,i} \cdot Y_{standardized,i})}{\sum_{i=1}^{23} (X_{standardized,i}^2)} = \frac{0,338}{23} = 0,015.$$

Осылайша Linear Regression моделі бойынша машиналық оқыту барысында анықталған, бидай өнімділігіне әсер етуші маңызды белгілердің ішінен ақпан айындағы (2-ай) жел бағыты әсерінің бета-коэффициенті ($\beta_1 = 0,015$) алынды (3-кесте). Бета-коэффициенттері машиналық оқытуға арналған бағдарламалау кітапханаларындағы арнайы функциялар көмегімен есептеледі. Машиналық оқытумен модельдер құру барысында бағдарламаны құру кезеңдері мен әр модель әдістері бойынша параметрлердің ерекшеліктері мен маңыздылықтарын ескеру қажет.

Төменде келтірілген 3-кестедегі бидай өнімділігіне әсер етуші климаттық факторлардың сандық мәндері сызықтық регрессия үшін β коэффициенттері болып табылады, олар (1) арқылы есептеледі. Ridge Regression үшін (2) арқылы айыппұл қосатын L2-регуляризация қолданылып есептелсе, ал Lasso Regression моделінде (3) арқылы есептелінетін L1-регуляризациясы қолданылады. Lasso-регрессиясы мен Ridge-регрессиясының артықшылықтарын қолдана отырып, Elastic Net регрессиясы L1 мен L2-регуляризациялары арасындағы қатынасты икемді түрде (4) реттеу арқылы есептейді. Қалған модельдер үшін машиналық оқыту барысында Python бағдарламалау тілінің sclearn кітапханасындағы әр модельдің арнайы кіріктірілген функциялары қолданылады.

Әзірленген модельдер бойынша бидай өнімділігіне әсер ететін маңызды факторлар туралы қорытынды жасауға мүмкіндік беретін бета-коэффициенттер мен белгілердің маңыздылығы көрсеткіштерінің салыстырмалы кестесі (3-кесте) төменде берілген.

Кесте 3 – Машиналық оқыту модельдері бойынша бидай өнімділігіне әсер етуші климаттық факторлардың бета-коэффициенттері мен белгілер маңыздылығының салыстырмалы кестесі

Модель атауы	Бидай өнімділігіне әсер етуші факторлар														
	Орташа температура Temp_average	Максималды темп. Temp_max	Минималды темп. Temp_min	Максимал. жел жылд. WindSpeed_max	Жел бағытының орташа мәні WindDir_average	Жауын-шашын (сомалық) Precipitation_sum	Жауын-шашын (орташа) Precipitation_average	Жылдық жауын- шашын (орташа) Annual_Precipitation	Минималды ылғалдық Hum_min	Максималды ылғалдық	Орташа ылғалдық Hum_average	Макс.топырақ темп. SoilTemp_max	Мин.топырақ темп. SoilTemp_min	Топырақ темп. орташа мәні SoilTemp_average	Ылғал.жетіспеу. орташа мәні DefHum_average
LinearRegression					0,015 (2-ай) -0,015 (10-ай) -0,013 (8-ай) -0,010 (1-ай) -0,011 (5-ай)	0,026 (11-ай) 0,024 (6-ай)		0,0345	-0,015 (2-ай) 0,013 (12-ай)						
RidgeRegression					-0,015 (10-ай) -0,014 (1-ай) 0,014 (2-ай) -0,018 (9-ай) -0,012 (8-ай)	0,025 (6-ай) 0,024 (11-ай)		0,038	-0,018 (2-ай) 0,012 (12-ай)						
LassoRegression					-0,018 (9-ай) 0,016 (12-ай) -0,012 (1-ай)	0,008 (12-ай) 0,084 (11-ай) -0,021 (10-ай) 0,014 (6-ай) 0,025 (8-ай)		0,035	-0,021 (2-мес)						
ElasticNetRegression	-0,304 (8-ай)			0,156 (8-ай)		0,263 (11-ай)			0,397 (8-ай) -0,243 (2-ай) 0,136 (6-ай)	0,572 (6-ай)	-0,145 (8-ай)	-0,227 (7-ай)	-0,171 (7-ай)		
DecisionTreeRegressor		0,186 (2-ай)			0,721 (9-ай)										0,774 (7-ай) 0,039 (5-ай)
RandomForestRegressor	0,057 (8-ай)	0,056 (7-ай)					0,029 (3-ай)		0,045 (8-ай)			0,027 (8-ай)	0,043 (7-ай) 0,038 (7-ай) 0,029 (5-ай)	0,071 (7-ай) 0,024 (8-ай)	
GradientBoostingRegressor	0,026 (8-ай)	0,021 (2-ай)	0,028 (8-ай)								0,026 (8-ай) 0,009 (5-ай)		0,055 (2-ай) 0,031 (8-ай)	0,690 (7-ай) 0,023 (5-ай) 0,010 (8-ай)	
ExtraTreesRegressor	0,043 (8-ай) 0,023 (2-ай)	0,034 (2-ай)			0,012 (5-ай)						0,036 (8-ай)		0,033 (2-ай) 0,013 (8-ай)	0,690 (7-ай) 0,019 (8-ай)	
Xgboost	0,422 (8-ай) 0,161 (2-ай) 0,004 (11-ай) 0,003 (6-ай)						0,0001 (3-ай)	0,002	0,0001 (8-ай)						0,407 (7-ай) 0,001 (1-ай)

● өте әлсіз байланыс 0-ден бастап 0,3-ке дейін ● әлсіз байланыс 0,3-тен бастап 0,5-ке дейін ● орташа байланыс 0,5-тен бастап 0,7-ге дейін ● жоғары байланыс 0,7-ден бастап 0,9-ға дейін ● өте жоғары байланыс 0,9-дан бастап 1,0-ге дейін

Нәтижесінде бидай өнімділігіне едәуір әсер етуші белгілер – шілдедегі ауа ылғалдылығының жетіспеушілігі 0,690-нан (Gradient Boosting Regressor шешім ағаштарының ансамблі мен Extra Trees Regressor) 0,774-ке дейін (Decision Tree Regressor), қыркүйек айындағы жел бағытының 0,721 көрсеткішімен (Decision Tree Regressor), маусымдағы максималды ылғалдылық 0,572 көрсеткішімен (Elastic Net Regression сызықтық әдісі) және тамыздағы ауа температурасы 0,422 көрсеткішімен (Xgboost) анықталды. Алынған нәтижелер бойынша, регрессиялық шешім ағашы (Decision Tree Regressor) моделі шілдедегі ауа ылғалдылығының тапшылығы (0,774) және қыркүйек айындағы желдің бағытының орташа мәнінің (0,721) бидай өнімділігі үшін ең маңызды факторлар болып табылатындығын дәлелдейді.

Decision Tree Regressor моделі бұны анық дәлелдейді және ең жоғарғы байланысты көрсетті. Шын мәнінде, Ақмола облысы бойынша шілдедегі ауа ылғалдылығының тапшылығы бидайдың вегетациялық кезеңінде едәуір ықпалын береді, сондай-ақ, қыркүйек айындағы

жел бағыты мен жылдамдығы бидай дәндерінің жерге құлауы есебінен оның өнімділігіне кері әсерін тигізеді. Егер жел қыркүйек айында болса және ол суық жел болған жағдайда, ылғалдың қайтарылуына әсері мол. Егер ыстық және құрғақ жел болса бұл тұқымдардың қатпарлығына әкелуі әбден мүмкін. Желдің жылдамдығы мен бағыты дақылды егу және химиялық өңдеу кезінде де маңызды орын алады. Әсер етуші фактор ретінде жауын-шашын, ылғалдылықтың мөлшері, ауа температурасының да бидай өнімділігіне тигізетін әсері мол. Жауын-шашынның жалпы мөлшерінің көрсеткіші бидай өнімділігі үшін маңызды климаттық факторлардың бірі болып саналады. Кейбір модельдерде есептеу барысында ескермеуімізге болмайтын және өте әлсіз байланысты көрсеткен факторлар да бар. Машиналық оқытумен болжамдық модельдерді құрудың келесі кезеңдерінде маңызды белгілер ретінде әлсіз байланыс көрсеткен кейбір белгілер деректер жиынынан тазартылады. Модельдердің болжамдарын келешекте кейбір қосымша сипаттамаларды және әсер етуші басқа да факторларды деректер жиынына қосу арқылы жақсартуға болады.

4. Қорытынды

Бұл зерттеуде Ақмола облысы, Ақкөл ауданының бидай өнімділігі бойынша деректер мен бидай өнімділігіне әсер етуші климаттық факторлардың метеодеректері негізінде құрылған модельдер нәтижесінде маңызды климаттық факторлар анықталды. Бидай өнімділігіне әсер етуші маңызды климаттық факторлардың сызықтық регрессия үшін бета-коэффициенттері және шешімдер ағашы мен бустинг үшін белгілер маңыздылығының көрсеткіштері ішінен ең жақсы нәтижелерді Decision Tree Regressor, Gradient Boosting Regressor, Extra Trees Regressor және Elastic Net Regression модельдері көрсетті. Decision Tree Regressor моделі маңызды фактор ретінде шілдедегі ылғалдылықтың жетіспеушілігі бойынша белгілер маңыздылығы көрсеткішінің ең жоғарғы 0,774 мәнін берсе, жел бағыты бойынша белгілер маңыздылығының көрсеткіші ретінде 0,721 мәнін берді. Gradient Boosting Regressor және Extra Trees Regressor модельдері шілдедегі ылғалдылықтың жетіспеушілігі бойынша белгілер маңыздылығы көрсеткішінің 0,690 орташа байланыс мәнін берді. Алынған нәтижелерден Elastic Net Regression моделі маусым айындағы максималды ылғалдылықтың коэффициенті ретінде 0,572 мәнін көрсетті. Сызықтық регрессияны, шешім ағаштарын және бустинг алгоритмдерін қоса алғанда, қарастырылған модельдердің ішінде орташа абсолютті пайыздық қателік (MAPE) көрсеткіші бойынша ең жақсы нәтижені Extra Trees Regressor моделі көрсетті. Extra Trees Regressor моделі бойынша MAPE 12,2% құрады, бұл басқа модельдерге қарағанда айтарлықтай төмен. Бұл климаттық факторларды ескере отырып, өнімділікті болжауда осы модельдің жоғары дәлдігін көрсетеді. Extra Trees Regressor моделінің жоғары дәлдігі оның сызықтық емес тәуелділіктер мен белгілер арасындағы өзара әрекеттесулерді тиімді өңдеу қабілетіне, сондай-ақ қайта оқытылуға төзімділігіне байланысты. Осы зерттеудің нәтижелері Extra Trees Regressor моделін пайдалану ауылшаруашылық жоспарлау мен шешім қабылдау үшін маңызды болып табылатын, өнімділікті болжау дәлдігін айтарлықтай жақсарту алатынын көрсетеді. Жазғы бидай өнімділігіне әсер ететін басқа да факторларға егжей-тегжейлі талдау жүргізу, сондай-ақ топырақ сапасы мен агротехникалық іс-шаралар сияқты қосымша деректерді қосу арқылы Extra Trees Regressor моделін одан әрі жақсарту жұмыстары алға қойылған. Алынған нәтижелер машиналық оқытумен болжамдық модельдер құрудың алғашқы маңызды міндеттері болып табылады және оқытудың келесі кезеңдеріне өтуге мүмкіндік береді.

Авторлар үлесі

Тәжібай Л.К. зерттеу жүргізу үшін деректерді қажетті форматқа өңдеді, бидай өнімділігіне әсер ететін маңызды факторларды анықтай отырып, болжамды деректерге талдау жүргізді және осы мақалада Linear Regression моделі үшін бета-коэффициенті мен орташа абсолютті пайыздық қатені анықтау бойынша есептеу мысалдарын келтірді, мақала мәтінін жазу, оның мазмұнын сыни тұрғыдан қайта қарау, жариялау үшін мақаланың соңғы нұсқасын бекіту бойынша жұмыстар жүргізді.

Мырзабекова Г. Е. бидай өнімділігін болжау үшін маңызды климаттық факторларды анықтаудың міндеттер қойылымын ұсынды және машиналық оқыту модельдерін әзірлеуге, мақала мәтінін жазу мен оның мазмұнын сыни тұрғыдан қайта қарауға үлес қосты.

Стыбаев Г. Ж. ауыл шаруашылығы ғылымдары саласындағы ғалым бола отырып, қажетті метеорологиялық деректерді ұсынды және алынған нәтижелерді түсіндіруге көмек көрсетті.

Муратова Г. К. деректерді алдын ала дайындауға, оның ішінде деректерді қалыпқа келтіруге үлес қосты.

Әдебиеттер тізімі

- 1 Кошкарлов А. В., Кошкарлова Т. А. Технологии сбора и анализа данных в цифровом сельском хозяйстве: барьеры и условия для внедрения и использования // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. – 2018. – №5. – С. 100-104.
- 2 Заяц О.А., Назарова Ю.Н., Стрижакова Е.А., Пенькова Р.И. Технологии Big Data в сельском хозяйстве // Fundamental research. – 2022. – № 7 – С. 35-40.
- 3 Xu X., Gao P., Zhu X., Guo W., Ding J., Li C., Wu X. Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China // Ecological Indicators. – 2019. – Vol. 101. – P. 943-953.
- 4 Alpaydin E. Introduction to Machine Learning. // The MIT Press. Cambridge.– 2014. 3rd edition. p. 3-4. [Electronic resource]. URL: [https://dl.matlabyar.com/siavash/ML/Book/Ethem%20Alpaydin-Introduction%20to%20Machine%20Learning-The%20MIT%20Press%20\(2014\).pdf](https://dl.matlabyar.com/siavash/ML/Book/Ethem%20Alpaydin-Introduction%20to%20Machine%20Learning-The%20MIT%20Press%20(2014).pdf) (Accessed: 20.02.2024).
- 5 Van K. T., Kassahun A., Catal C. Crop yield prediction using machine learning: A systematic literature review // Computers and Electronics in Agriculture. – 2020. - Vol. 177. – P. 105709. DOI: <https://doi.org/10.1016/j.compag.2020.105709>.
- 6 Китов В. Практические аспекты машинного обучения // Открытые системы. СУБД. – 2016 №1 [Электронды ресурс]. URL: <https://www.osp.ru/os/2016/01/13048648> (жүгінген күн: 20.02.2024).
- 7 Тәжібай Л.К., Мырзабекова Г.Е. Ауыл шаруашылығындағы Data Mining технологиялары // М.А.Гендельманның 110 жылдығына арналған «Сейфуллин оқулары-19» халықаралық ғылыми-практикалық конференция материалдары. –2023. –Т.І, Б.ІІІ. –137-141 б.
- 8 Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture // Environmental Research Letters – 2018. – Vol.13, №11. – P. 1748-9326. DOI: <https://dx.doi.org/10.1088/1748-9326/aae159>.
- 9 Elbasi E., Zaki C., Topcu AE., Abdelbaki W., Zreikat AI., Cina E., Shdefat A., Saker L. Crop Prediction Model Using Machine Learning Algorithms // Applied Sciences. – 2023. – 13(16) – P. 9288. DOI: <https://doi.org/10.3390/app13169288>.
- 10 Gajula A., Singamsetty J., Dodda V., Kuruguntla L. Prediction of crop and yield in agriculture using machine learning technique // 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). – 2021. – P.1-5. DOI: 10.1109/ICCCNT51525.2021.9579843.
- 11 Tussupov J., Yessenova M., Abdikerimova G., Aimbetov A., Baktybekov K., Murzabekova G., Aitimova U. Analysis of Formal Concepts for Verification of Pests and Diseases of Crops Using Machine Learning Methods // IEEE Access – 2024. – Vol.12. – P. 19902–19910.
- 12 Yessenova M., Abdikerimova G., Ayazbaev T., Ismailova A., Beldeubayeva Z., Ainagulova A., Mukhanova A. The effectiveness of methods and algorithms for detecting and isolating factors that negatively affect the growth of crops // International Journal of Electrical and Computer Engineering. —2003. – Vol. 13, № 2. – P. 1669-1679. DOI: 10.11591/ijece.v13i2.
- 13 Куценогий П.К., Каличкин В.К., Пакуль А.Л., Куценогий С.П. Машинное обучение в качестве инструмента прогнозирования урожайности сельскохозяйственных культур// Российская сельскохозяйственная наука. – 2021. – № 1. – С. 72-75. DOI: 10.31857/S2500262721010178
- 14 Архипова М. Ю. Моделирование урожайности зерновых культур сельскохозяйственных регионов с использованием технологий компьютерного зрения // Экономика региона. – 2022. – Т. 18, вып. 2. – С. 581-594. DOI: <https://doi.org/10.17059/ekon.reg.2022-2-20>.
- 15 Акшалов К.А. Высокоэффективные технологии возделывания зерновых культур в полусухих условиях Северного Казахстана //Методические рекомендации. – 2009. – 57 С.
- 16 Erenstein O., Jaleta M., Abdul Mottaleb K., Sonder K., Donovan J., Braun H.-J. Global Trends in Wheat Production, Consumption and Trade // Wheat Improvement: Food Security in a Changing Climate. –2022. –P. 47 – 66.
- 17 Zachow M., Nóia Júnior, Rogério de S., Asseng S. Seasonal climate models for national wheat yield forecasts in Brazil // Agricultural and Forest Meteorology. – 2023. – Vol. 342. – P.109753-109767.

Л.К. Тажибай¹, Г.Е. Мурзабекова², Г.Ж. Стыбаев³, Г.К. Муратова⁴
Казахский агротехнический исследовательский университет имени С. Сейфуллина,
пр. Женис 62, г. Астана, Казахстан

Модель определения климатических факторов, влияющих на урожайность пшеницы, с помощью машинного обучения

Аннотация. Определения важных климатических факторов для прогнозирования урожайности пшеницы позволяет разработать эффективные стратегии адаптации для растениеводства в сельском хозяйстве. Это помогает минимизировать негативные последствия климатических изменений в производстве пшеницы. Целью данного исследования является разработка моделей машинного обучения, таких, как линейные модели, деревья решений и бустинговые модели, для определения климатических факторов, влияющих на урожайность пшеницы. Исследование основано на данных по урожайности пшеницы и метеоданных Аккольского района Акмолинской области за 2000-2022 годы, включая температуру воздуха, влажность, осадки, скорость и направление ветра, температуру поверхности почвы и дефицита влажности воздуха. В этой статье для разработки прогностических моделей были использованы шесть линейных моделей машинного обучения Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, SVR, Gaussian Process Regressor, четыре модели деревьев решений Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor и две модели бустинга LGBM регрессор, Xgboost. В результате были выявлены признаки, существенно влияющие на урожайность пшеницы, среди которых дефицит влажности воздуха в июле, максимальная влажность в июне и температура воздуха в августе. Лучшие результаты показали модели деревьев решений. Полученные данные были интерпретированы и могут использоваться при принятии хозяйственных решений.

Ключевые слова: машинное обучение, прогнозная модель, важность признаков, интеллектуальный анализ, принятие решений.

L.K. Tazhibay¹, G.Y. Murzabekova², G. Zh. Stybayev³, G.K. Muratova⁴
Seifullin Kazakh Agrotechnical Research University, Zhenis ave. 62, Astana, Kazakhstan

Model for determining climatic factors affecting wheat productivity by machine learning

Abstract. Determining important climatic factors for predicting wheat yield allows the development of effective adaptation strategies for crop production in agriculture. This helps minimize the negative impacts of climate change on wheat production. The objective of this research is to develop machine learning models such as linear models, decision trees and boosting models to identify climate factors affecting wheat yield. The study is based on wheat yield data and weather data from the Akkol district of the Akkola region for the period 2000-2022. This includes data on air temperature, humidity, precipitation, wind speed and direction, soil surface temperature and air humidity deficit. In this article, six linear machine learning models were used to develop predictive models: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, SVR, Gaussian Process Regressor, four decision tree models Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor and two boosting models LGBM regressor, Xgboost. As a result, signs were identified that significantly affect wheat yield. These findings included a deficit of air humidity in July, maximum humidity in June and air temperature in August. Decision tree models showed the best results. The data obtained were interpreted and can be used to inform business decisions in agricultural structures.

Key words: machine learning, predictive model, feature importance, predictive analysis, decision making.

References

- 1 Koshkarov A. V., Koshkarova T. A. Tekhnologii sbora i analiza dannyh v cifrovom sel'skom hozyajstve: bar'ery i usloviya dlya vnedreniya i ispol'zovaniya. [Koshkarov A.V., Koshkarova T.A. Technologies for collecting and analyzing data in digital agriculture: barriers and conditions for implementation and use], *Sovremennaya nauka: aktual'nye problemy teorii i praktiki. Seriya: Estestvennye i Tekhnicheskie Nauki* [Modern science: current problems of theory and practice. Series: Natural and Technical Sciences]. 2018. №05. P. 100-104.
- 2 Zayac O.A., Nazarova YU.N., Strizhakova E.A., Pen'kova R.I. Tekhnologii Big Data v sel'skom hozyajstve. [Zayats O.A., Nazarova Yu.N., Strizhakova E.A., Penkova R.I. Big Data technologies in agriculture], *Fundamental research*. 2022. №7. P. 35-40.
- 3 Xu X., Gao P., Zhu X., Guo W., Ding J., Li C., Wu X. Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China, *Ecological Indicators*. 2019. Vol. 101. P. 943-953.
- 4 Alpaydin E. Introduction to Machine Learning, The MIT Press. Cambridge. 2014. 3rd edition. p. 3-4. [Electronic resource]. URL: [https://dl.matlabyar.com/siavash/ML/Book/Ethem%20Alpaydin-Introduction%20to%20Machine%20Learning-The%20MIT%20Press%20\(2014\).pdf](https://dl.matlabyar.com/siavash/ML/Book/Ethem%20Alpaydin-Introduction%20to%20Machine%20Learning-The%20MIT%20Press%20(2014).pdf) (Accessed: 20.02.2024).
- 5 Van K. T., Kassahun A., Catal C. Crop yield prediction using machine learning: A systematic literature review, *Computers and Electronics in Agriculture*. 2020. Vol. 177. P. 105709. DOI: <https://doi.org/10.1016/j.compag.2020.105709>.
- 6 Kitov V. Prakticheskie aspekty mashinnogo obucheniya [Kitov V. Practical aspects of machine learning], *Otkrytye sistemy. SUBD [Open systems. DBMS]*. №01. 2016. [Electronic resource]. URL: (<https://www.osp.ru/os/2016/01/13048648>).

- 7 Tazhibaj L.K., Myrzabekova G.E. Auysharuashylygyndagy Data Mining tekhnologiyalary [Tazhibaj L.K., Myrzabekova G.E. Data Mining technologies in agriculture], M.A. Gendel'mannyn 110 zhyldygyna arnalgan «Sejfullin okulary-19» halykaralyk gylymi-praktikalyk konferenciya materialdary [Proceedings of the international scientific and practical conference "Seifullin readings-19" dedicated to the 110th anniversary of M.A. Handelman]. 2023. T.I, B.III. 137-141 b.
- 8 Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture, Environmental Research Letters 2018. Vol.13, №11. P. 1748-9326. DOI: <https://dx.doi.org/10.1088/1748-9326/aae159>.
- 9 Elbasi E., Zaki C., Topcu AE., Abdelbaki W., Zreikat AI., Cina E., Shdefat A., Saker L. Crop Prediction Model Using Machine Learning Algorithms, Applied Sciences. 2023. Vol. 13, №16. P. 9288. DOI: <https://doi.org/10.3390/app13169288>.
- 10 Gajula A., Singamsetty J., Dodda V., Kuruguntla L. Prediction of crop and yield in agriculture using machine learning technique, 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). 2021. P.1-5. DOI: 10.1109/ICCCNT51525.2021.9579843.
- 11 Tussupov J., Yessenova M., Abdikerimova G., Aimbetov A., Baktybekov K., Murzabekova G., Aitimova U. Analysis of Formal Concepts for Verification of Pests and Diseases of Crops Using Machine Learning Methods, IEEE Access. 2024. Vol.12. P. 19902-19910.
- 12 Yessenova M., Abdikerimova G., Ayazbaev T., Ismailova A., Beldeubayeva Z., Ainagulova A., Mukhanova A. The effectiveness of methods and algorithms for detecting and isolating factors that negatively affect the growth of crops, International Journal of Electrical and Computer Engineering. 2003. Vol. 13, №2. P. 1669-1679. DOI: 10.11591/ijece.v13i2.
- 13 Kucenogij P.K., Kalichkin V.K., Pakul' A.L., Kucenogij S.P. Mashinnoe obuchenie v kachestve instrumenta prognozirovaniya urozhajnosti sel'skohozyajstvennyh kul'tur. [Kutsenogiy P.K., Kalichkin V.K., Pakul A.L., Kutsenogiy S.P. Machine learning as a tool for crop yield forecasting], Rossijskaya sel'skohozyajstvennaya nauka. [Russian agricultural science]. 2021.№1. P. 72-75. DOI: 10.31857/S2500262721010178
- 14 Arhipova M. YU. Modelirovanie urozhajnosti zernovyh kul'tur sel'skohozyajstvennyh regionov s ispol'zovaniem tekhnologij komp'yuternogo zreniya. [Arkipova M. Yu. Modeling the yield of grain crops in agricultural regions using computer vision technologies], Ekonomika regiona.[Economy of the region]. 2022. Vol. Is. 18, vyp. 2. P. 581-594. DOI: <https://doi.org/10.17059/ekon.reg.2022-2-20>.
- 15 Akshalov K.A. Vysokoeffektivnyye tekhnologii vozdeleyvaniya zernovyh kul'tur v poluzasushlivykh usloviyah Severnogo Kazakhstana. [Akshalov K.A. Highly efficient technologies for cultivating grain crops in semi-arid conditions of Northern Kazakhstan], Metodicheskie rekomendacii. [Guidelines]. 2009. 57p.
- 16 Erenstein O., Jaleta M., Abdul Mottaleb K., Sonder K., Donovan J., Braun H.-J. Global Trends in Wheat Production, Consumption and Trade, Wheat Improvement: Food Security in a Changing Climate. 2022. P. 47 – 66.
- 17 Zachow M., N6ia J6nior, Rog6rio de S., Asseng S. Seasonal climate models for national wheat yield forecasts in Brazil, Agricultural and Forest Meteorology. 2023. Vol. 342. P.109753-109767.

Авторлар туралы мәліметтер:

Тәжібай Ләззат Каналбайқызы – байланыс үшін автор, докторант, С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Жеңіс даңғылы 62, Астана, Қазақстан.

Мырзабекова Гүлден Есләмбекқызы – ф.-м.ғ.к., доцент, С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Жеңіс даңғылы 62, Астана, Қазақстан.

Стыбаев Ғани Жасымбекұлы – а.-ш.ғ.к., профессор, С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Жеңіс даңғылы 62, Астана, Қазақстан.

Муратова Гульжан Клычовна – т.ғ.к., қауымдастырылған профессор м.а., С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Жеңіс даңғылы 62, Астана, Қазақстан

Information about authors:

Tazhibay Lazzat Kanalbaykizi –corresponding author, PhD student, S. Seifullin Kazakh Agrotechnical Research University, Zhenis ave. 62, Astana, Kazakhstan.

Murzabekova Gulden Yeslyambekovna – Candidate of Physical and Mathematical Sciences, Associate Professor, S. Seifullin Kazakh Agrotechnical Research University, Zhenis ave. 62, Astana, Kazakhstan.

Stybayev Gani Zhasymbekovich - Candidate of Agricultural Sciences, Professor, S. Seifullin Kazakh Agrotechnical Research University, Zhenis ave. 62, Astana, Kazakhstan.

Muratova Gulzhan Klychovna – Candidate of Technical Sciences, acting Associate Professor, S. Seifullin Kazakh Agrotechnical Research University, Zhenis ave. 62, Astana, Kazakhstan.

*Қабылданған күні: 29.03.2024. Өңдеуден кейін: 29.06.2024.
Мақұлданған күні: 29.06.2024. Онлайн қолжетімді: 30.06.2024.*