

Н.К. Мукажанов¹, Ж.М. Алибиева¹, Л.Ш. Черикбаева²,
А.М. Касенхан¹, А.С. Еримбетова¹

¹Satbayev University, Алматы, Қазақстан

²al-Farabi Kazakh National University, Алматы, Қазақстан

E-mail: mukazhan@mail.ru

ЕГУГЕ САЙ КЕЛЕТІН ДАҚЫЛДАРДЫ БОЛЖАУ ҮШІН АНСАМБЛДІК ОҚЫТУ ҮЛГІСІН ӘЗІРЛЕУ

Аңдатпа. Ауыл шаруашылығы халықты азық-түлікпен қамтамасыз етудің негізгі саласы және еліміздің экономикасында маңызды рөл атқарады. Қазақстан Республикасында 25,7 млн. гектар егістік жерлер бар. Оны тиімді пайдалану жолдарының бірі заманауи технологиялардың мүмкіндіктерін пайдалану, соның ішінде жасанды интеллект бағыты машиналық оқыту. Ауыл шаруашылығында машиналық оқыту мүмкіндіктерін пайдалану тиімділікті арттыру, процестерді оңтайландыру, таңдалған жерге сәйкес дақылды анықтау, егін шығымдылығын болжау, өсімдіктерді зиянкестерден қоғау және т.б. фермерлер үшін маңызды тапсырмаларды шешуге көмектеседі. Бұл мақалада бірнеше машиналық оқыту әдістерін ансамблге біріктіру арқылы белгілі бір жердің топырақ құрамы мен жауын-шашын мөлшеріне байланысты сол жерге егуге сәйкес келетін дақылды болжау тапсырмасын шешу қарастырылды. Ансамблдік үлгіні құру үшін 10-астам машиналық оқытудың әдістеріне әртүрлі тестер жүргізілді. Тест нәтижесінде жоғарғы дәлдік көрсеткен шешім ағашы классификаторы (Decision tree classifier), логистикалық регрессия (Logistic regression), Naive Bayes классификаторы, К-ең жақын көрші (K-Nearest Neighbors) және тереңнен оқытудың бірнеше қабатты нейрондық желі (Multi-layer perceptron classifier) әдістері стекинг тәсілі бойынша ансамблік үлгіге біріктіріліп жаңа екі ансамблдік үлгілер (Stacking with ML models және Stacking with ML and NN models) алынды. Құрылған ансамблдік үлгілердің дәлдік көрсеткіштері сәйкесінше Stacking with ML models - 99.95%, ал Stacking with ML and NN models - 76.9% болды.

Түйінді сөздер. Машиналық оқыту, ансамблдік оқыту, деректер жиыны, ауыл шаруашылық дақылдары, стекинг, болжау.

Кіріспе.

Ақпараттық технология саласында басты тредтердің бірі жасанды интеллект екені белгілі. Жасанды интеллект құру және оның мүмкіндіктерін қолданысқа енгізуде әлем бойынша зерттеулер жүргізіліп, түрлі аппараттық-программалық шешімдер ұсынылуда. Жасанды интеллект пен цифрландыруды тиімі пайдалануды қажет ететін саланың бірі ауыл шаруашылығы. Компьютерлік көруді, машиналық оқытуды және IoT қолданбаларын пайдалану өнімділікті арттыруға, сапаны жақсартуға және сайып келгенде фермерлер мен онымен байланысты салалардың табыстылығын арттыруға көмектеседі. Егін жинаудың жалпы өнімін арттыру үшін ауыл шаруашылығы әлемінде дәл егіншілік маңызды болып табылады [1]. Мысалы, ақылды суару жүйелері, дақылдардың ауруын болжау, дақылдарды таңдау, ауа-райын болжау және ең төменгі қолдау бағасын анықтау ауыл шаруашылығында қолданылатын әдістердің мысалдары болып табылады. Бұл әдістер фермерлерден аз жұмысты талап ете отырып, егістік өнімділігін арттырады [2]. Дақылдардың шығымдылығын бағалау әр түрлі мақсаттарда қолданылуы мүмкін, соның ішінде фермерлерге өнім көлемін арттыруға көмектесу, тыңайтқыштарға және басқа да

ауылшаруашылық өнімдеріне сұраныс пен ұсыныс циклін оңтайландыру, бағаларды болжау және ауыл шаруашылығын сақтандыру бойынша тәуекел деңгейін есептеу [3].

Машиналық оқытуды қолдану арқылы деректерді толық талдай отырып, фермерге жақсы өнім алуды, тиімді дақылды егуді ұсынуға, егін шығымдылығын болжауға, зиянкестермен күресуде қолдануға болады [4, 5, 6]. Жақсы өнім алу үшін топырақтың түрі мен құнарлылығын, сонымен қатар жауын-шашын мөлшерін ескеру қажет, егер ол құрғақ болса, құрғақшылыққа төзімді дақыл түрін таңдаған дұрыс, ал сулы-батпақты жер болса, соған бейім егін түріне барған дұрыс. Таңдалған жерге сай келетін дақыл түрін машиналық оқыту арқылы дәлірек болжауға болады. Бұл тапсырма бойынша жүргізілген зерттеулердің көпшілігі [7,8,9] жеке-жеке оқыту үлгілеріне назар аударған. Бірнеше машиналық оқыту әдістерін ансамблдік үлгіге біріктіру арқылы болжам дәліктерін жоғарылатуға болады. Ұсынылып отырған зерттеу барысында ансамблдік оқыту көмегімен топырақтың құрамы мен ауа-райы жағдайларын ескере отырып егуге болатын дақыл түрін анықтау қарастырылады.

Зерттеудің мақсаты климаттық жағдайлар мен топырақ құрамы қасиеттеріне байланысты, белгіленген жер параметрлеріне сәйкес егуге келетін дақылдарды болжау үшін ансамблді оқыту моделін әзірлеу және бағалау. Бұл мақала фермерлерге дақылдарды таңдау және жерді басқару бойынша негізделген шешім қабылдау құралын ұсыну арқылы тұрақты ауыл шаруашылығы мақсатына ықпал етеді. Аумаққа сәйкес келетін дақылды дұрыс таңдау бір жағынан шығындарды азайтса, екінші жағынан қолданылатын тыңайтқыштар көлемін азайту арқылы қоршаған ортаға жағымды әсер етеді.

Материалдар мен тәсілдер

Белгілі бір аумаққа егілетін дақылдың сәйкестігі мен егін шығымдылығын деректерді талдау, машиналық оқыту, үлкен деректерді сараптау, сенсорлық технологияларды қолдану арқылы болжау бойынша көптеген зерттеулер жүргізілген. Болжамдар жасауға қажетті деректерде түрлі жолдармен жиналған, мысалы, сенсорлар арқылы жинау, аумақ бойынша алдын-ала жинақталған ауыл-шаруашылық статистикалық деректері, data mining үшін дайындалған тестілік деректер қоймасы, машиналық оқытуға арналған тестілік деректер жиыны.

[4] мақалада авторлар сәйкес дақыл мен егін шығымдылығын болжау бойынша жазылған мақалаларға шолу жасаған. Негізгі жұмыс өзге авторлардың қарастырған жердің параметрлері мен классификация/кластеризация алгоритмдеріне салыстырмалы талдауларға бағытталған. Зерттеу әдебиеттік шолу мен талдауға арналғандықтан, эксперименттік тексерулер берілмеген. Талдаулар нәтижелерін қорытындылай келе авторлар, таңдалған аумаққа сәйкес егін түрін анықтау үшін топырақ құрамы мен климаттық жағдай параметрлері негізгі екенін көрсеткен.

[10] жұмыс авторлары қоршаған орта мен топырақ ерекшеліктеріне негізделген дақылдарды болжайтын жүйені ұсынған. Жүйе болжамдарды қоршаған орта сипаттамалары атрибуттары және топырақ сипаттамалары атрибуттары арқылы табады. Жиналған деректер жинағы алдын ала өңделеді және орауыш мүмкіндігін таңдау әдістерін қолдану арқылы алдын ала өңделген деректерден маңызды атрибуттар шығарылды. Содан кейін олар белгілі бір аумақта өсіруге қолайлы дақылды табу үшін өндірілген атрибуттар бойынша классификация алгоритмін қолданады.

[11] авторлар дақылдарды таңдауда дәстүрлі тәсіл агроном мамандардың тәжірибесіне және заманауи машиналық оқытуға сүйенуге болатынын айта келе, машиналық оқыту жолдарын қолдануды перспективалы бағыт ретінде ұсынған. Жұмыста машиналық оқыту әдістерінің ішінде k-nearest neighbor, Naive Bayes, Regression, Random Forest қолдануды зерттеген. Ансамблдік оқытуды жекелеп қарастырмаған.

Ансамблдік оқыту көмегімен сәйкес дақылды болжау бойынша жасалынған жұмыстар [12,13,14]. [12] жұмыс авторлары ағашқа негізделген ансамблдік оқыту жетілдіруге арналған. Нәтижесінде әдістің дәлдігін 99,32 % дейін жеткізген. Бұл өте жақсы көрсеткіш. [13] жұмыста Red Fox Optimization және Recurrent Neural Network (RNN) ансамблінің комбинациясын қолданатын егін ұсынысы мен кірісті болжау үлгісін берген. Зерттеу нәтижелері соңғы әдістермен салыстырғанда ұсынылған модельдің жоғары өнімділігін көрсетті, бұл әртүрлі өнімділік көрсеткіштерімен дәлелденген. [14] мақалада кездейсоқ орман және логистикалық регрессия алгоритмдерін пайдалана отырып, егін өнімділігін болжау және тыңайтқыштарды ұсыну жүйесін ұсынды. Ұсынылған шешімде егін өнімділігін болжау және тыңайтқышты ұсыну арқылы ауыл шаруашылығын оңтайландыру үшін машиналық оқытудың екі алгоритмін біріктірген. Жүйені әзірлеу үшін олар Random forest және логистикалық регрессияны қолданған. Бұл модель жоғарыда аталған мәселелерді шешуге және кірісті арттыруға болатын гибриді ML тәсілдерінің мысалы ретінде қызмет етеді.

Қарастырылып отырған тапсырмада интернет заттарды (IoT) қолдану [15] жұмыста берілген. Авторлар фермаларға IoT сенсорларды орнату арқылы ауа-райын бақылау, температураны өлшеу, топырақ құрамын анықтау арқылы егуге болатын егін түрін болжауды қарастырған. Сенсорлардан алынған деректерді талдау негізінде болжамдар ұсынады, ал екінші болжамдар бойынша өзі шешім қабылдайды.

[16] жұмыста авторлар статистикалық үлгілер арқылы және MARS егін шығымдылығын болжау жүйесінде Қазақстанда бидайдың шығымдылығын болжау бойынша жазған. Бұдан басқа көптеген шет елдік зерттеулерді көруге болады және олар нақты ауыл шаруашылығы, егін шығымдылығын болжауда түрлі жолдарды қарастырған [17-21].

Отандық ғалымдардың жұмыстарының ішінде, [22] мақалада нақты ауыл шаруашылығында отандық және шет елдік жетістіктерді енгізу бойынша талдаулар жасап, дәнді дақылдар шығымдылығы бойынша болжамдар жасау бағытында зерттеулер мен бірқатар тәжірибелер жасаған. [23] зерттеу жұмысы температура мен спутниктік снимоктар арқылы Солтүстік Қазақстанда бидай шығымын модельдеуге арналған.

Жоғарыда жасалынған әдебиеттерге шолудан, машиналық оқыту әдістері егін шаруашылығында көптеген тапсырмаларды шешетіні көруге болады. Тапсырмалардың ішінен белгілі бір аумақ бойынша егуге болатын/сәйкес егін түрін болжау тапсырмасын қарастырамыз. Қазақстан территориясындағы топырақ құрамы, климаттық жағдайы, су мөлшері аймақтарға байланысты әр түрлі. Әрбір агроклиматтық аймақ белгілі бір дақылдарды өсіре алады. Осыған сүйене отырып, біз фермерге белгілі бір климаттық аймаққа қолайлы дақыл түрін ұсынуымыз керек. Сондықтан біз климаттық аймаққа негізделген қолайлы дақылды таңдауды жақсырақ түсіну үшін машиналық оқыту мен ансамблдік оқытуды пайдаланамыз. Белгілі бір аумақ үшін дақыл таңдауына әсер ететін кейбір негізгі белгілер: топырақ факторлары, климаттық факторлар, судың болуы (су көздерінің жеткіліктілігі) және т.б. Сәйкес дақылды анықтау үшін бастықы деректер ретінде kaggle сайтынан ауылшаруашылық дақылдарын ұсыну деректер жиыны [24] пайдаланылады. Бұл деректер жиынында көптеген дақылдар Қазақстанның климаттық жағдайына мүлдем сәйкес келмейді, сондықтан оңтайландыру мақсатына тек біздің елемізде өсірілетін дақылдар алынады. Деректер жиынына әр түрлі өңдеулер мен түрлендірулер жасалынады. Әрі қарай, машиналық оқыту үлгілері құрылады және оқытылады. Оқытылған үлгілердің ішінен жоғары дәлдік көрсеткен үлгілер стекинг әдісі бойынша ансамблдік үлгіге біріктіріледі. Осылай, үздік машиналық оқыту әдістерін біріктіре отырып, ансамбальдік оқытудың үлгісін әзірленеді. Машиналық оқыту мен ансамблдік оқыту әдістері тестіленіп, олардың дәлік көрсеткіштеріне салыстырмалы талдаулар жасалынады.

Деректерді алдын ала өңдеу. Жерге егілетін дақылды анықтау үшін топырақтың құрамы N, P, K, pH көрсеткіштері мен климаттық жағдай көрсеткіштері қолданылады. Деректер жиыны келесідегідей параметрлерден тұрады:

Деректер өрісі:

N - топырақтағы азот мөлшерінің коэффициенті

P - топырақтағы фосфор мөлшерінің коэффициенті

K - топырақтағы калий мөлшерінің коэффициенті

temperature - цельсий бойынша температура

humidity - салыстырмалы ылғалдылық %

ph - топырақтың pH мәні

rainfall - жауын-шашын мөлшері мм

label – ауыл шаруашылығы дақылының атауы

Қарастырылып отырған деректер жиынынан егуге болатын дақылдар ретінде күріш (rice), жүгері (maize), бұршақ (chickpea), бүйрек бұршақтары (kidneybeans), жүзім (grapes), қарбыз (watermelon), қауын (muskmelon), алма (apple), мақта (cotton) және жасымық (lentil) алынды.

```
PATH = 'dataset/dataset.csv'  
df = pd.read_csv(PATH)
```

```
df.head()
```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

1 сурет - Талдауға арналған dataset

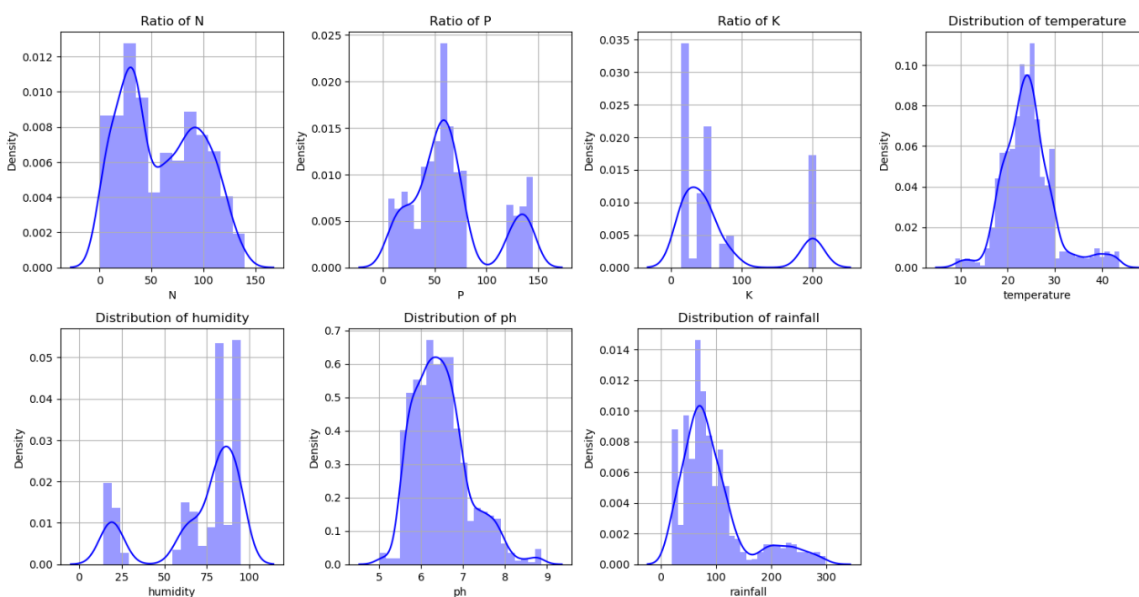
Зерттеуде пайдаланылатын деректер жинағының параметрлер бойынша қысқаша статистикалық көрсеткіштері 2-суретте берілген. Статистикалық дереткерден азот деңгейі 0,00-ден 140-қа дейін, фосфор деңгейі 5-тен 145-ке дейін және калий деңгейлері 15-тен 205-ке дейін мәндер қабылдағанын байқауға болады. Бұлар топырақ құрамы бойынша параметрлер. Келесі маңызды параметрлер климаттық көрсеткіштер, температура 8,8 ден 41,9 дейін, ылғалдылық 14,2-ден 94,9 дейін, ал жауын-шашын мөлшері 20,2-ден 298,5 дейінгі мәндерді қабылдаған. Жоғарыда айтылғандай, біз бұл деректер жиынында тек Қазақстан аумағында егілетін ауыл-шаруашылық өсімдіктерді ғана алдық. Бұл параметр мәндеріне сәйкес келетін Қазақстанда климаттық аймақтар бар.

```
df.describe()
```

	N	P	K	temperature	humidity	ph	rainfall
count	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000
mean	59.87500	64.74200	69.89000	23.429983	68.219091	6.440782	89.074813
std	38.97962	39.288663	67.718385	4.414797	26.228809	0.679440	57.409876
min	0.00000	5.00000	15.00000	8.825675	14.258040	5.005307	20.211267
25%	26.00000	38.00000	21.00000	20.624560	61.859357	5.939346	53.143849
50%	60.00000	58.00000	45.00000	23.538031	80.625750	6.323270	74.169169
75%	95.00000	77.00000	79.00000	25.836515	85.019314	6.771959	104.935828
max	140.00000	145.00000	205.00000	41.948657	94.962187	8.868741	298.560117

2 сурет - Деректер жиынының статистикалық көрсеткіштері.

Деректер жиындағы мәндерді визуалды барлау үшін `displot` диаграммасы қолданылды. `Displot` диаграммаларында пішін, тенденциялар, таралу және ауытқулар сияқты негізгі сипаттамаларды бөліп көрсететін деректердің көрнекі қысқаша мазмұнын беруге болады. Деректер қалыпты үлестірімге сәйкес келетінін, шектен шыққанын немесе басқа да өзгешеліктердің болуын көруге болады. 3-суретте деректер жиынының N, P, K, temperature, humidity, ph және rainfall атрибуттары бойынша `displot` диаграммасы берілген.

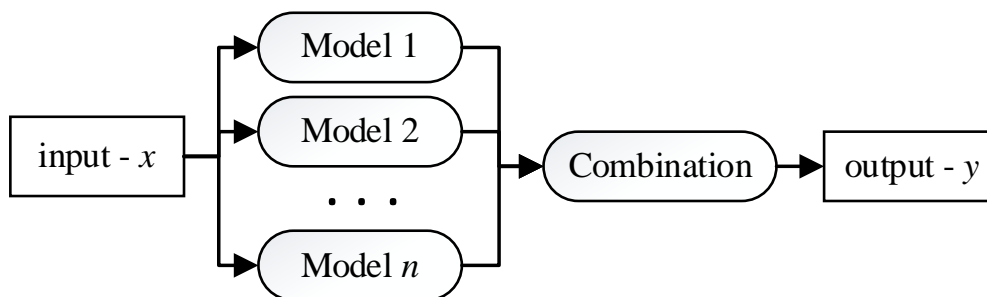


3 сурет - Деректер жиынындағы атрибуттар бойынша `displot` диаграммасы

`Displot` диаграммада көрсетілгендей параметрлер мәндері бір-бірінен өзгеше таратылған. Бұл атрибуттарда әр түрлі өлшем көрсеткіштері мен мәндер пайдаланылуына байланысты.

Деректер жиыны дайындалғаннан кейін, деректер жиынын оқыту және тестілеу деректеріне бөлу орындалады. Деректер жиынын оқыту және тест деректеріне бөлу `scikit-learn` кітапханасының `train_test_split()` функциясы арқылы жүзеге асырылды.

Сәйкес келетін дақылды болжауда k-nearest neighbor, Naive Bayes, Regression, Random Forest, Recurrent Neural Network және т.б. машиналық оқыту және нейрондық желі әдістері қолданылатындығын әдебиеттерге шолуда айтқан болатынбыз. Біз әрі қарай машиналық оқытудың әртүрлі үлгілерін біріктіріп, сәйкес дақылды болжаудың ансамблдік оқытуға негізделген жаңа әдісін ұсынатын боламыз. Жалпы ансамблдік оқыту әдістері біртекті (homogeneous) және біртекті емес (heterogeneous) болып екі типке бөлінеді. Біздің қарастыратын стекинг әдісі біртекті емес болып табылады. Ансамблдік оқыту үлгілерінің жалпылама архитектурасы 4-суретте берілген [25,26,27].

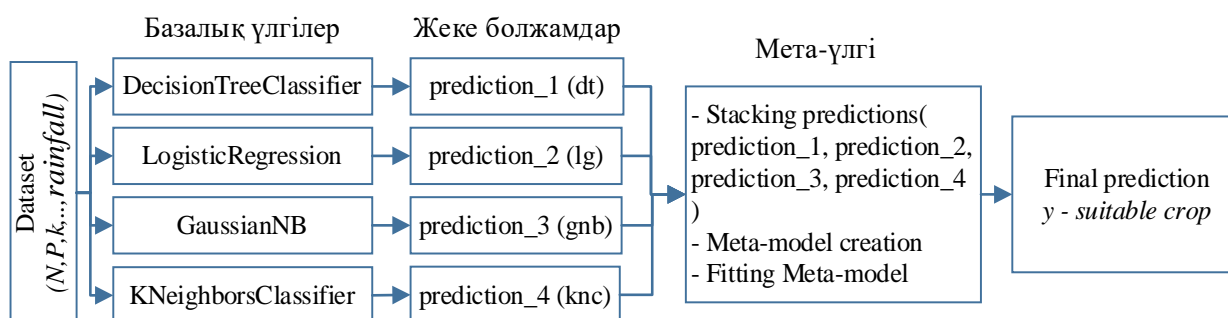


4 сурет - Ансамблдік оқытудың жалпылама архитектурасы

Стекинг-бұл соңғы қорытынды болжамды алу үшін бірінші деңгейлі модельдер немесе базалық үлгілер деп аталатын көптеген негізгі модельдердің болжамдарын біріктіретін машиналық оқыту стратегиясы [25]. Бұл бір деректер жинағына көптеген негізгі модельдерді оқытуды, содан кейін олардың болжамдарын соңғы болжамды алу үшін мета-үлгіге немесе екінші деңгейлі үлгі деп аталатын жоғары деңгейлі модельге жіберуді орындайды. Үлгілерді комбинациялаудың негізгі идеясы бір модельді қолданғаннан гөрі жақсы болжау өнімділігін алу үшін әртүрлі базалық үлгілердің болжамдарын біріктіру. Жоғары дәлдікті ансамблдік үлгіні алу үшін стекинг арқылы базалық үлгілер ретінде тек машиналық оқыту әдістерін біріктіру және машиналық оқыту әдістерін нейрондық желі арқылы классификациялау әдісімен біріктіру нұсқалары құрылды.

Нәтижелер.

Бірінші, тек машиналық оқыту әдістерін комбинациялау арқылы ансамблдік үлгіні әзірлеу жүзеге асырылды. Машиналық оқыту үлгілерін біріктіру арқылы бірнеше ансамблдік үлгілер құрылды. Олардың ішінде дәлдік көрсеткіші жоғары болған үлгі (Stacking with ML models) осы жұмыста ұсынылды. Stacking with ML models үлгісінің архитектурасы 5-суретте берілген.



5 сурет - Stacking with ML models ансамблдік үлгі архитектурасы

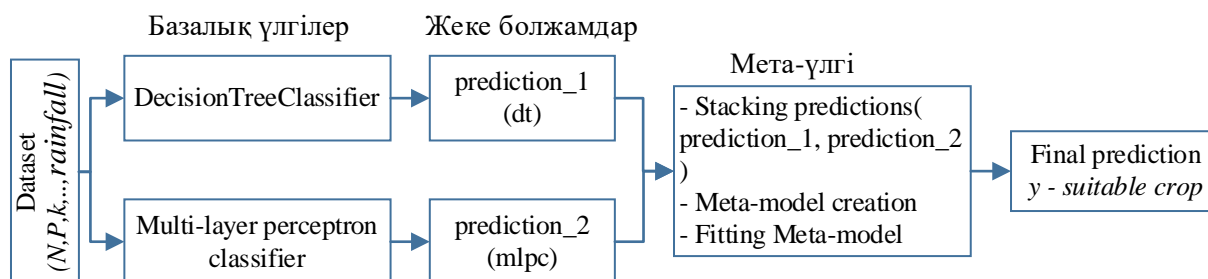
5-суретте көріп отырғандарыңыздай, шешім ағашы классификаторы (Decision tree classifier), логистикалық регрессия (Logistic regression), Naive Bayes классификаторы, К-ең жақын көрші (K-Nearest Neighbors) әдісі базалық үлгілер ретінде алынып, олардың болжамдары мета-үлгіге біріктірілді. Мета-үлгі әр түрлі үлгілерден алынған болжамдарды біріктіре отырып, қорытынды болжамды береді. Әзірленген үлгіге тест деректер бойынша жасалған сынық нәтижелері 6-суретте берілген. Суртте көрсетілгендей үлгінің дәлдік көрсеткіші 99,95 пайызға тең.

Stacking_with_ML_models_Accuracy: 99.9513661112118799

	precision	recall	f1-score
0	1.00	1.00	1.00
1	1.00	1.00	1.00
2	1.00	0.96	1.00
.			
accuracy			0.99
macro avg	0.97	0.98	1.00
weighted avg	1.00	0.92	0.99

6 сурет - Машиналық оқыту әдістерін біріктіру арқылы әзірленген ансамблдік үлгінің дәлдік көрсеткіштері

Екінші, қарапайым машиналық оқыту үлгілерін бірнеше қабатты нейрондық желі (Multi-layer perceptron classifier) үлгісімен біріктірілген ансамбл үлгісі алынды (Stacking with ML and NN models). Stacking with ML and NN models үлгісінің архитектурасы 7-суретте берілген.



7 сурет- Stacking with ML and NN models ансамблдік үлгі архитектурасы

Stacking with ML and NN үлгісінің дәлдігі 76,9 пайызға тең. Басқада, precision, recall және f1-score көрсеткіштері толығырақ 8-суретте берілген.

Stacking_with_ML_and_NN_Accuracy: 76.9090909090909

	precision	recall	f1-score
.			
accuracy			0.77
macro avg	0.66	0.72	0.66
weighted avg	0.69	0.77	0.69

8 сурет - Stacking with ML and NN әдісінің дәлдік көрсеткіштері

Stacking with ML and NN үлгісі берген 76,9 пайыз басқа үлгілермен салыстырғанда (1-кесте) жақсы көрсеткіш деп айтуға келмейді. Машиналық оқыту мен нейрондық желі әдістерін ансамблге біріктіру арқылы әзірленген үлгілердің дәлдігін жақсартуды келесі жұмыстарда қарастыратын боламыз.

Машиналық оқыту әдістерін біріктіру арқылы әзірленген ансамблдік үлгі сәйкес дақылды анықтауда қолдану мысалы 9-суретте берілген. Бірінші, таңдалған жердің топырақ қасиеттерінің параметр мәндері мен климаттық жағдай параметрлерінің мәндері ($X = [N, P, K, \text{temperature}, \text{humidity}, \text{ph}, \text{rainfall}]$) беріледі. Ондан кейін оқытылған машиналық оқыту үлгілерінен жеке болжамдар алынып, олар бір матрицаға біріктіріледі ($X_{\text{meta}} = [\text{prediction1}, \text{prediction2}, \text{prediction3}, \text{prediction4}]$). Біріктірілген матрица мета-үлгіге параметр ретінде беріледі де, мета-үлгі қорытынды болжамды (y) қайтарады.

```
# сәйкес келетін дақылды болжау
N, P, k, temperature, humidity, ph, rainfall = 40, 20, 55, 25, 30, 35, 30
X_new_val = np.array([[N, P, k, temperature, humidity, ph, rainfall]])
dt_1_pred = dt_1.predict(X_new_val)
lg_1_pred = lg_1.predict(X_new_val)
gnb_1_pred = gnb_1.predict(X_new_val)
knc_1_pred = knc_1.predict(X_new_val)

# базалық үлгілердің болжамдарын бір матрицаға белгілермен біріктіру
X_meta = np.column_stack((dt_1_pred, lg_1_pred, gnb_1_pred, knc_1_pred))

# Мета-үлгіні пайдаланып болжам жасау
crop_dict = {'Күріш': 0, 'Жүгері': 1, 'Бұршақ': 2, 'Фасоль (үрме бұршақ)': 3, 'Жасымық': 4,
             'Жүзім': 5, 'Қарбыз': 6, 'Қауын': 7, 'Алма': 8, 'Мақта': 9}
y_new_pred = meta_model_1.predict(X_meta)
for key, val in crop_dict.items():
    if y_new_pred == val:
        print("Өсіруге сәйкес келетін дақыл - ", key)
```

Өсіруге сәйкес келетін дақыл - Қарбыз

9 сурет - Stacking with ML models ансамблдік үлгіні сәйкес дақылды анықтау үшін қолдану

9-суретте ансамблдік үлгінің программалық жүзеге асырылуы жасалған. Суреттің төменгі жағында, болжам бойынша «қарбыз» дақылы берілген параметрлерге ең жақсы сәйкес келетін, өсіруге болатынын көрсетіп тұр. Ұсынылып отырған ансамблдік үлгінің программалық әзірлемесі қолданбалы қосымшаларға қолдануға дайын.

Талқылау.

Бұл мақалада ұсынылып отырған стекинг әдісі бойынша әзірленген Stacking with ML models пен Stacking with ML and NN models ансамблдік үлгілерінің дәлдік көрсеткіштері сәйкесінше 99,95% және 76,9% тең. Тестілік сынақтар жүргізу барысында, осыған дейін бустинг және бэгинг әдістері бойынша әзірленген Gradient Boosting, Extra Trees, AdaBoost, Bagging және Random Forest ансамблдік үлгілеріне де сықтар жүргізілді (нәтижелер 1-кестеде берілген). Бұл үлгілердің ішінде өті жақсы нәтиже берген Gradient Boosting, Bagging және Random Forest үлгілері. Осы жұмыста қарастырылған барлық ансамблдік оқыту үлгілерінің дәлдік көрсеткіштерін 1-кестеден көруге болады.

1 кесте - Ансамблдік оқыту үлгілерінің көрсеткіштері

Алгоритмдер	Дәлдік, % (accuracy)	weighted avg		
		precision	recall	f1-score
Gradient Boosting	98.18	0.98	0.98	0.98
Extra Trees	88.63	0.90	0.89	0.89
AdaBoost	9.54	0.05	0.10	0.06
Bagging	98.86	0.99	0.99	0.99
Random Forest	99.31	0.99	0.99	0.99
Stacking with ML models	99.95	1.00	0.92	0.99
Stacking with ML and NN models	76.9	0.69	0.77	0.69

Кестеден көріп отырғандарыңыздай Random Forest үлгісі 99.31 пайыз дәлдік берді, шамамен біздің әзірлеген Stacking with ML models үлгісі дәлдігімен шамалас. Сонымен қатар, Bagging және Random Forest үлгілерінде сәйкес дақылды анықтау тапсырмасында қолдануға болады. Ал, AdaBoost үлгісінің дәлдігі мүлдем төмен, 9.54 пайыз, біздің тапсырмада қолдануға ұсынуға келмейді. Сондай-ақ біздің ұсынған Stacking with ML and NN models үлгісінің дәлдігі (76.9%) жақсы емес. Нақты қолданбаларда пайдалану үшін жетілдіруді қажет етеді.

Қорытынды.

Бұл зерттеу жұмысында, басты назарды таңдалған жерге сәйкес келетін дақыл түрін болжауда машиналық оқыту әдістерін, соның ішінде ансамблдік оқыту әдістерін қолдануды қарастырдық. Деректер жиіні ретінде kaggle сайтынан ауылшаруашылық дақылдарын ұсыну деректер жиіні [24] пайдаланылды. Тәуелсіз параметрлер ретінде топырақтағы азот мөлшері (N), топырақтағы фосфор мөлшері (P), топырақтағы калий мөлшері (K), цельсий бойынша температура, салыстырмалы ылғалдылық, топырақтың қышқылдығы (pH) және жауын-шашын мөлшері алынды. Ал, мақсатты параметр ретінде дақыл атауы қарастырылды. Бұл деректер жиінін алдын ала өңдеуде тек Қазақстанның климаттық жағдайында өсіруге келетін өсімдіктер қалдырылды, категориялық деректер сандық түрге ауыстырылды және параметр мәндерін бір қылып келтіру бойынша түрлендірулер жасалынды. Зерттеу нәтижесі ретінде келесі теориялық қорытындылар мен практикалық нәтижелер алынды:

1) Стекинг әдісі бойынша бірнеше машиналық оқыту әдістерін комбинациялау арқылы ансамблдік үлгісі (Stacking with ML models) әзірленді және шешім ағашы мен бірнеше қабатты нейрондық желі (Multi-layer perceptron classifier) үлгісін біріктіріп екінші ансамбл үлгісі (stacking with ML and NN models) құрылды.

2) Әзірленген үлгілер оқытылып, дәлдік көрсеткіштері анықталынды және дәлдіктері сәйкесінше 99,95% және 76,9% тең. Сонымен қатар, stacking with ML models ансамблдік үлгісі программалық жүзеге асырылып, болжамды ұсыну мысалы берілді. Бұл әзірлеме сәйкес дақылды анықтау қосымшаларында пайдалануға дайын.

3) Фермерлерге экономикалық жағынан тиімді және түрлі тәуелділіктерді алдын алу үшін таңдалған жерге егуге нақты сәйкес келетін дақылды анықтау үшін жасанды интеллекті мүмкіндіктерін пайдалану қажет. Келешекте, ауылшаруашылық саласында ақпараттық технологияны, соның ішінде жасанды интеллект қолдану ауықымы арта бермек. Бұл өз кезегінде қазіргі уақытта қолмен есептеп, анықтайтын көптеген жұмыстарды алмастыратыны сөзсіз.

ӘДЕБИЕТТЕР

[1] Rishi G., Akhilesh K. Sh., Oorja Ga., Krishna M., Shahreen K., Zirawani B., Hairulnizam M. and Salama A. M., “WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics”, IEEE Access, 2021.

[2] Rakesh Kumar, Singh, Prabhat Kumar / “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 2015:138-145.

[3] Ajinkya Patil, Pavan Patil, Shrikant Kokate, “Crop Prediction System using Machine Learning Algorithms”, International Journal of Advancements in Engineering and Technology (IJAET). 2020;1(1).

- [4] I. L. Shahana, V. P. Nijth Kamal, R. L. Anoop, Ancy Francis / Crop Prediction Methods: A Comparative Study / Asian Journal of Agricultural Extension, Economics & Sociology / 40(8): 228-236, 2022; Article no. AJAEES.84828 ISSN: 2320-7027 /
- [5] Sita Rani, Amit Kumar Mishra, Aman Kataria, Saurav Mallik 4, Hong Qin / Machine learning-based optimal crop selection system in smart agriculture / Scientific Reports / (2023) 13:15997 / <https://doi.org/10.1038/s41598-023-42356-y>
- [6] Kiran S., Priyanka R. S., Pooja N., Tushar S. L., Mayuri S. J. / Crop Prediction and Fertilizer Recommendation Using Machine Learning /. International Journal of Engineering Research and Applications www.ijera.com / ISSN: 2248-9622, Vol. 13, Issue 3, March 2023, pp. 28-32
- [7] S. Bhanumathi, M. Vineeth, N. Rohit, Crop yield prediction and efficient use of fertilizers, in: 2019 International Conference on Communication and Signal Processing, ICCSP, 2019, pp. 0769–0773, <http://dx.doi.org/10.1109/ICCSP.2019.8698087>.
- [8] M. Rashid, B.S. Bari, Y. Yusup, M.A. Kamaruddin, N. Khan, A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction, IEEE Access 9 (2021) 63406–63439, <http://dx.doi.org/10.1109/ACCESS.2021.3075159>.
- [9] Y.J.N. Kumar, V. Spandana, V.S. Vaishnavi, K. Neha, V.G.R.R. Devi, Supervised machine learning approach for crop yield prediction in agriculture sector, in: 2020 5th International Conference on Communication and Electronics Systems, ICCES, 2020, pp. 736–741, <http://dx.doi.org/10.1109/ICCES48766.2020.9137868>
- [10] A. Suruliandi, G. Mariammal / Crop prediction based on soil and environmental characteristics using feature selection techniques / Mathematical and Computer Modelling of Dynamical Systems Methods, Tools and Applications in Engineering and Related Sciences Volume 27, 2021 - Issue 1 / <https://doi.org/10.1080/13873954.2021.1882505>
- [11] PRITESH P., PRANAV A., MANAS B., SIDDHI T., ADITYA M. / Crop Selection and Yield Prediction using Machine Learning Approach / Current Agriculture Research Journal / ISSN: 2347-4688, Vol. 11, No. (3) 2023, pg. 968-980
- [12] Isaac K. N., Adib Z., Owusu N.-B., Adebayo F.A., Frimpong K. / A predictive analytics model for crop suitability and productivity with tree-based ensemble learning / Decision Analytics Journal 8 (2023) 100311 / <https://doi.org/10.1016/j.dajour.2023.100311>
- [13] P.S.S. Gopi, M. Karthikeyan, Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model / Multimedia Tools Appl. (2023) <http://dx.doi.org/10.1007/s11042-023-16113-2>.
- [14] K.P.K. Devan, B. Swetha, P. Uma Sruthi, S. Varshini, Crop yield prediction and fertilizer recommendation system using hybrid machine learning algorithms, / IEEE 12th International Conference on Communication Systems and Network Technologies, CSNT, 2023, pp. 171–175, <http://dx.doi.org/10.1109/CSNT57126.2023.10134662>.
- [15] Fan-Hsun Tseng Hsin-Hung Cho / Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis / IEEE Access PP (99):1-1 / August 2019 DOI: 10.1109/ACCESS.2019.2935564
- [16] Paula Romanovska, Bernhard Schauburger, Christoph Gornott. Wheat yields in Kazakhstan can successfully be forecasted using a statistical crop model / European Journal of Agronomy Volume 147, July 2023, 126843
- [17] S. Chouhan, D. Singh, and A. Singh, An improved feature selection and classification using decision tree for crop datasets, International Journal of Computer Applications 142 (2016), pp. 5–8. doi:<https://doi.org/10.5120/ijca2016909966>. [Crossref], [Google Scholar]
- [18] D. Heckmann, U. Schlüter, and A.P.M. Weber, Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra, Mol Plant 10 (2017), pp.

878–890. doi:<https://doi.org/10.1016/j.molp.2017.04.009>. [Crossref] [PubMed] [Web of Science ®], [Google Scholar]

[19] P.S. Maya Gopal and R. Bhargavi., Feature selection for yield prediction in Boruta algorithm, International Journal of Pure and Applied Mathematics 118 (2018), pp. 139–144. [Google Scholar]

[20] A. Bahl, B. Hellack, M. Balas, A. Dinischiotu, M. Wiemanne, J. Brinkmann, A. Luch, B.Y. Renard, and A. Haase, Recursive feature elimination in random forest classification supports nanomaterial grouping, NanoImpact 15 (2019), pp. 100179. doi:<https://doi.org/10.1016/j.impact.2019.100179>. [Crossref] [Web of Science ®], [Google Scholar]

[21] P.S. Maya Gopal and R. Bhargavi., Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms, Appl. Artif. Intell. 33 (2019), pp. 621–642. doi:<https://doi.org/10.1080/08839514.2019.1592343>. [Taylor & Francis Online] [Web of Science ®], [Google Scholar]

[22] Marzhan Sadenova, Nail Alikuly Beisekenov, Petar Sabev Varbanov, Ting Pan / Application of Machine Learning and Neural Networks to Predict the Yield of Cereals, Legumes, Oilseeds and Forage Crops in Kazakhstan / MDPI journal: Agriculture 2023, 13(6), 1195; <https://doi.org/10.3390/agriculture13061195>

[23] Bekbayeva A. ASSESSMENT OF USING LAND SURFACE TEMPERATURE (LST) AND SATELLITE REMOTE SENSING FOR WHEAT YIELD MODELLING IN THE NORTH KAZAKHSTAN REGION / С.Сейфуллин атындағы Қазақ агротехникалық университетінің Ғылым жаршысы (пәнаралық) = Вестник науки Казахского агротехнического университета им. С.Сейфуллина (междисциплинарный). - 2019. - №4 (103). - С.33-42

[24] Kaggle-сайты. Ауыл шарушылық деректері бойынша деректер жиыны / <https://www.kaggle.com/datasets/varshitanalluri/crop-recommendation-dataset> / (деректерді алу уақыты - 20.05.2024)

[25] Alok Kumar, Mayank Jain / Ensemble Learning for AI Developers / book / Business Media New York, ISBN-13 (pbk): 978-1-4842-5939-9 ISBN-13 (electronic): 978-1-4842-5940-5, <https://doi.org/10.1007/978-1-4842-5940-5>

[26] Babak Asadi, Ramez Hajj / Prediction of asphalt binder elastic recovery using tree-based ensemble bagging and boosting models / Construction and Building Materials Volume 410, 5 January 2024, 134154 / <https://doi.org/10.1016/j.conbuildmat.2023.134154>

[27] Aleksandra P., Michael A., Galina M. / Heterogeneous versus Homogeneous Machine Learning Ensembles / Heterogeneous versus Homogeneous Machine Learning Ensembles, 2015 / DOI: 10.1515/itms-2015-0021

Нуржан Мукажанов, PhD, Satbayev University, Алматы, Қазақстан, mukazhan@mail.ru

Жибек Алибиева, PhD, Satbayev University, Алматы, Қазақстан, alibievajibek@gmail.com

Ляйля Черикбаева, PhD, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, cherikbayeva.Lyailya@gmail.com

Арай Касенхан, PhD, Satbayev University, Алматы, Қазақстан, a.kassenkhan@satbayev.university

Айгерим Еримбетова, PhD, Satbayev University, Алматы, Қазақстан, aigerian@mail.ru

РАЗРБОТКА МОДЕЛИ АНСАМБЛЕГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПОДХОДЯЩИХ КУЛЬТУР ДЛЯ ПОСЕВА

Аннотация. По всему миру сельское хозяйство является основной отраслью продовольственного обеспечения населения и играет важную роль в экономике нашей страны. В Республике Казахстан имеется 25,7 млн гектаров пахотных земель. Одним из способов его эффективного использования является использование возможностей современных технологий, в том числе направлении искусственного интеллекта - машинное обучение. Использование возможностей машинного обучения в сельском хозяйстве может способствовать повышению эффективности, оптимизации процессов, определению подходящих культур для выбранного место, прогнозированию урожайности, защите растений от вредителей и т. д. помогает решать важные для фермеров задачи. В этой статье было рассмотрено решение задачи прогнозирования подходящих культур для посева, в зависимости от состава почвы и количества осадков, путем объединения нескольких методов машинного обучения в ансамбль. Для создания ансамблевой модели были проведены различные тесты более 10 методов машинного обучения. В результате теста методы дерева решений, логистическая регрессия, классификатор Наивного Байеса, классификатор K-Nearest Neighbours и многослойный перцептрон показали более высокую точность. Эти методы были объединены в ансамблевую модель с помощью стекинга и были получены две новые ансамблевые модели Stacking with ML models и Stacking with ML and NN models. Показатели точности ансамблевых моделей составили соответственно Stacking with ML models - 99,95%, Stacking with ML and NN models - 76,9%.

Ключевые слова. Машинное обучение, ансамблевое обучение, наборы данных, сельскохозяйственные культуры, стекинг, прогнозирование.

Nurzhan Mukazhanov, PhD, Satbayev University, Almaty, Kazakhstan, mukazhan@mail.ru

Zhibek Alibiyeva, PhD, Satbayev University, Almaty, Kazakhstan, alibievajibek@gmail.com

Lyailya Cherikbayeva, PhD, al-Farabi Kazakh National University, Almaty, Kazakhstan, Cherikbayeva.Lyailya@gmail.com

Arai Kasenkhan, PhD, Satbayev University, Almaty, Kazakhstan, e-mail: a.kassenkhan@satbayev.university;

Aigerim Yerimbetova, PhD, Satbayev University, Almaty, Kazakhstan, aigerian@mail.ru

DEVELOPMENT OF AN ENSEMBLE LEARNING MODEL FOR PREDICTING SUITABLE CROPS FOR PLANTING

Abstract. All over the world, agriculture is the main sector of food supply for the population and plays an important role in the economy of our country. The Republic of Kazakhstan has 25.7 million hectares of arable land. One of the ways to use it effectively is to use the capabilities of modern technologies, including the direction of artificial intelligence - machine learning. Using machine learning capabilities in agriculture can help improve efficiency, optimize processes, identify suitable crops for a given location, predict yields, protect plants from pests, etc., helping to solve problems important to farmers. This article examined how to solve the problem of predicting suitable crops to plant based on soil composition and rainfall by combining several machine learning methods into an ensemble. To create the

ensemble model, various tests were carried out on more than 10 machine learning methods. As a result of the test, decision tree methods, logistic regression, Naive Bayes classifier, K-Nearest Neighbors classifier and multilayer perceptron showed higher accuracy. These methods were combined into an ensemble model using stacking and two new ensemble models were obtained: Stacking with ML models and Stacking with ML and NN models. The accuracy indicators of the ensemble models were, respectively, Stacking with ML models - 99.95%, Stacking with ML and NN models - 76.9%.

Keywords. Machine learning, ensemble learning, datasets, agricultural crops, stacking, forecasting.

Редакцияға түсті / Поступила в редакцию / Received 13.08.2024
Жариялауға қабылданды / Принята к публикации / Accepted 27.01.2025