

«М.А. Гендельманның 110 жылдығына арналған «Сейфуллин оқулары – 19» халықаралық ғылыми-практикалық конференциясының материалдары = Материалы международной научно-практической конференции «Сейфуллинские чтения – 19», посвященной 110 - летию М.А. Гендельмана» - 2023.- Т.І, Ч.ІІІ.- Б. 137-141.

ӘОЖ 004.942

АУЫЛ ШАРУАШЫЛЫҒЫНДАҒЫ DATA MINING ТЕХНОЛОГИЯЛАРЫ

*Тәжібай Л.К., 2-курс докторанты,
Мырзабекова Г.Е., ф.-м.ғ.к., доцент
«С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті»
КеАҚ, Астана қ.*

Бүгінгі күні өсімдік шаруашылығының немесе мал шаруашылығының толық циклін бақылауға, ауыл шаруашылығындағы шаруашылық шешімдерді талдауға, болжауға және оңтайландыруға мүмкіндік беретін заманауи цифрлық технологияларды қолданусыз ауыл шаруашылығы өндірісін тиімді басқаруды елестету мүмкін емес.

Жер халқы санының өсу процестері шеңберінде ауыл шаруашылығының тиімділігін арттыру және өнімнің экологиялық тазалығына қойылатын талаптарды арттыру көптеген елдердің аграрлық секторының алдында тұрған маңызды міндет болып табылады. 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. Деректерді зияткерлік талдаудың әртүрлі әдістерін қолдану агробизнестің тиімділігін оның дамуының әртүрлі кезеңдерінде арттыруға мүмкіндік береді. Бұл ауыл шаруашылығы индустриясында, әсіресе, өсімдік шаруашылығы саласында өзекті. Деректерді зияткерлік талдау негізінде алқаптардың күйін бақылауға, компьютер немесе смартфон экранында талдау нәтижелерін тексеру арқылы мәселелерді анықтауға және болжауға болады.

Арнайы бағдарламалармен үлкен деректерді талдау ауыл шаруашылығының әр түрлі мәселелерін шешуге мүмкіндік береді: өнімділікті болжау, егісті отырғызу немесе жинау үшін қолайлы кезеңді анықтау, тыңайтқыштар схемасын есептеу және көптеген тағы басқалары [2]. Осы саладағы әртүрлі ғалымдардың зерттеулері, деректерді зияткерлік талдау сияқты технологияларды қолданудың жұмыс тиімділігінің жоғарылауына, шығындарды азайтуға және кәсіпорын үшін тәуекелдерді азайтуға әкелетін шешім қабылдауға оң әсер ететінін растайды.

Үлкен деректердің көмегімен әртүрлі модельдер құруға, заңдылықтарды анықтауға және болашақта процестердің беталысын болжауға болады. Болжамды модельдер тек сауда мен маркетингте ғана емес, сонымен қатар кез-келген басқа саладағы трендтерді анықтауға қызмет ете алады. Мысалы, ауыл шаруашылығында болжамды модельдер өнімділік немесе құрғақшылық, қабылданған шешімдердің тиімділігі туралы болжам жасау үшін қолданылады.

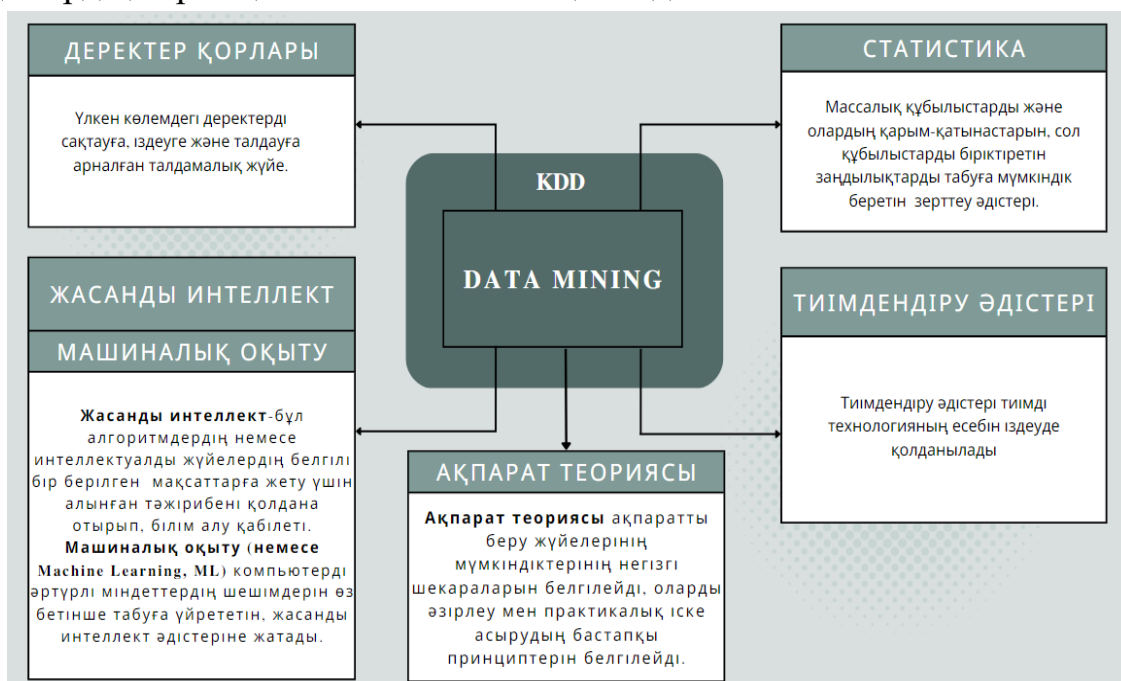
Үлкен деректерді талдаудың келесідей бірнеше әдістері бар: деректерді алдын ала өңдеу немесе Data Mining, машиналық оқыту (ML), нейрондық желілер, болжамдық талдау, статистикалық талдау. Data mining-бұл деректер жиындарынан білім алуға байланысты мәселелердің барлық саласын қамтитын ақпараттық технологиялар бағыты. Бұл деректер жиынтығындағы ақпаратты анықтау процесін білдіреді; сонымен қатар, деректер базасындағы білімді анықтау (Knowledge Discovery in Databases, KDD) ретінде де белгілі. Data mining арқасында екі нәтижеге қол жеткізуге болады: қолда бар деректерді сипаттау немесе болашаққа болжам жасау.

Деректерді зияткерлік талдау негізінде оңтайландыру, генетикалық алгоритмдер, бейнені тану, статистика, Data Mining және тағы басқалары сияқты математикалық әдістерді белсенді қолдану, сондай-ақ ақпараттың визуалды көрінісін пайдалану идеясы жатыр.

"Зияткерлік талдау" және "Data Mining" терминологияларында пәндер шеңберінде шешілетін мәселелер көлемінде айырмашылық бар. Негізінен, Data Mining әдістері деректерді талдау процестеріне және модельдеуге бағытталады, ал деректерді зияткерлік талдау деректер базасынан білім алу процесіне қатысты мәселелердің барлық спектрін қарастырады.

Деректердегі пайдалы білімді іздеу (KDD) процесінде деректерді пайдалы ақпаратқа аудару кезінде деректерді, таңдау критерийлерін және нәтижелердің интерпретацияларын дайындау қажет. Ол үшін көптеген математикалық әдістер, соның ішінде, статистикалық әдістер қолданылады.

Төмендегі 1-суретте білім салаларының өзара байланысы және олардың Data Mining-ке қатынасы ұсынылған, бұл ретте, деректерді зияткерлік талдау әдістердің барлық жиынтығын өзіне қамтиды.



Сурет 1. Білім салаларының байланысы және олардың Data Mining-ке қатынасы.

Өнімділікті болжау дәлме-дәл егіншіліктің күрделі міндеттерінің бірі болып табылады және бүгінгі күнге дейін көптеген модельдер, соның ішінде, машиналық оқыту модельдері ұсынылды және расталды. Бұл мәселе бірнеше деректер жиынтығын пайдалануды талап етеді, өйткені өнімділік климат, ауа-райы, топырақ, тыңайтқыштарды пайдалану және тұқым сорттары сияқты көптеген әртүрлі факторларға байланысты [3].

Біздің зерттеуімізде Ақмола облысы, Ақкөл ауданы бойынша "Қазгидромет" РМК сайтынан алынған метеодеректер (<https://www.kazhydromet.kz/ru/>) және Ұлттық статистика бюросының сайтынан (<https://www.stat.gov.kz/>) алынған бидай өнімділігі бойынша 2000-2022 жылдар аралығындағы деректер қолданылды. Бидай өнімділігін талдау және болжау үшін машиналық оқыту әдістері, оның ішінде, сызықтық алгоритмдер, шешімдер ағаштары алгоритмдері және градиентті бустинг қолданылды.

ML 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 жоғары өнімді 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 әдістерін ендіру міндетті қоюды, деректерді зерттеуді және оларды дайындауды, модельдеуді (болжамды модель құру және оның параметрлерін таңдау), алынған модельді бағалауды, ендіруді және қолдауды қамтиды. 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].

Болжау кезінде машиналық оқытуды қолдану келесі кезеңдерден тұрады: міндеттерді қою; деректерді дайындау; әдісті таңдау; модельді нақтылау; модельдердің композициясы; деректерді шектеу; машиналық оқыту шектері; сапа критерийі; ендіру және қолдау.

Зерттеу жұмысында машиналық оқытумен болжамдық модель құру процесі бірнеше кезеңдермен жүзеге асырылады. Алдымен, қойылған міндеттер анықталып алынғаннан соң, деректердің жинағы таңдалынып, синтезделеді. Деректер жинағын әзірлеу барысында бидай өнімділігіне әсер етуші негізгі факторлар зерттелді. Болжамдық модель құрудың келесі кезеңдерінде, деректерді тазалау және қалыпқа келтіру бойынша жұмыстардан соң, модельді таңдау маңызды рөл атқарады. Жұмыста маңызды деректерді таңдау кезеңінде бірнеше регрессиялық, шешімдер ағаштары модельдері және регрессиялық әдіс пен нейрондық желілерге баламалы градиентті бустинг модельдері артықшылықтарына қарай таңдалынып, қолданылды. Модельдеу процесі машиналық оқытуда қолданылатын Python бағдарламалау тілінің кітапханаларымен жүзеге



асырылды. Болжамдық модельдерге салыстырмалы талдау жүргізіліп, бидай өнімділігіне әсер етуші факторлардың маңыздылығы айқындалды. Болжамды модельді машиналық оқытумен құру процесінің сызбалық көрінісі 2-суретте көрсетілген.

Сурет 2. Машиналық оқытуды қолданумен болжамды модель құру процесі.

Солтүстік Қазақстан өңіріндегі Ақмола облысы, Ақкөл ауылының жаздық бидай өнімділігін болжау моделін әзірлеу міндеті бойынша және деректер жинағы ретінде бидай өнімділігіне әсер етуші факторлардың деректері зерттеліп, дайындалды.

Жаздық бидай ұзақ күндізгі дақылға және қысқа вегетациялық кезеңге жатады. Жаздық бидайдың көшеттері кешеуілдеген көктемнің аязына минус 8 - минус 10°C шамасында оңай шыдайды, керісінше, піскен кезде бидай күздің ерте аязымен гүлдену кезеңінде минус 1 - минус 2°C - та , сүтті пісу кезеңінде - минус 2-минус 4°C-та оңай зақымдалады [7]. Тұқымның өнуі үшін қолайлы жағдайлар 12-15°C температурада және топырақтың жоғарғы қабатындағы топырақтың ылғалдылығы 70-90% ең аз далалық ылғал сыйымдылығында жасалады.

Жаздық бидай ерекшелігіне қарай, модельдің деректер жиынына өнімділікке әсер етуші факторлар ретінде, ауа температурасы, жылдық жауын-шашын көлемі, топырақ температурасы, салыстырмалы ылғалдылық, қанықтылық тапшылығы, желдің бағыты мен жылдамдығы бойынша метеодеректері алынды.

Деректер жиынтығы машиналық оқыту алгоритмдеріне сәйкес форматқа түрлендірілді. Жүргізілген зерттеуде өңделген деректер алдын-ала екі бөлікке (оқыту және тестілеу), атап айтқанда, оқытуға арналған деректердің 70%-ына және тестілеуге арналған деректердің 30%-ына бөлінген.

Біздің зерттеуімізде өнімділікті болжау барысында маңызды деректерді талдау үшін келесідей: LinearRegression; RidgeRegression; LassoRegression; ElasticNetRegression; SVR; GaussianProcessRegressor; DecisionTreeRegressor; RandomForestRegressor; GradientBoostingRegressor; ExtraTreesRegressor; LGBMRegressor; xgboost машиналық оқыту модельдері қолданылды. Жақсы нәтиже алу үшін оқытылған машиналық модельдерге салыстыру жұмыстары жүргізіле отырып, талданды.

Жоғарыда аталған модельдер Python бағдарламалау тілінің pandas, numpy, sklearn, seaborn, matplotlib, т.б. әртүрлі кітапханаларын қолданумен әзірленіп, арнайы интерфейстік ортада нәтижелері шығарылды.

Бұл зерттеу жұмысында бидай өнімділігін болжау үшін әртүрлі машиналық оқыту алгоритмдерін қолдану барысында LinearRegression, GradientBoostingRegressor, xgboost модельдері бидай өнімділігіне әсер етуші факторлары әсерінің салыстырмалы түрде жақсы көрсеткіштерін көрсетті. Модельдердің болжамдарын келешекте кейбір қосымша сипаттамаларды және әсер етуші басқа да факторларды деректер жиынына қосу арқылы жақсартуға болады.

Пайдаланған әдебиеттер тізімі

- 1 Кошкарров А. В., Кошкаррова Т. А. Технологии сбора и анализа данных в цифровом сельском хозяйстве: барьеры и условия для внедрения и использования [Мәтін] // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. -2018. -№05. - С. 100-104
- 2 Заяц О.А., Назарова Ю.Н., Стрижакова Е.А., Пенькова Р.И. Технологии Big Data в сельском хозяйстве [Мәтін] // 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 [Text]: a case study in Jiangsu Province, China. // Ecol. 2019. Ind. 101, p. 943–953. (<https://doi.org/10.1016/j.ecolind.2019.01.059>.)
- 4 Alpaydin E. Introduction to Machine Learning. 2nd ed. [Text]. // The MIT Press. Cambridge. 2010. (https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf) (жүгінген күн: 20.02.2023)

- 5 Van K. T., Kassahun A., Catal C. Crop yield prediction using machine learning: A systematic literature review. [Text]. Computers and Electronics in Agriculture. 2020. Volume 177, p. 105709 (<https://doi.org/10.1016/j.compag.2020.105709>)
- 6 Китов В. Практические аспекты машинного обучения [Мәтін]. // Открытые системы. СУБД» выпуск №01.- 2016.- (<https://www.osp.ru/os/2016/01/13048648>)
- 7 Акшалов К.А. Высокоэффективные технологии возделывания зерновых культур в полужасушливых условиях Северного Казахстана. [Мәтін] // Методические рекомендации. - 2009. -57 с.
- 8 Janmejay P., Pant R.P., Manoj K. S., Devesh P. S., Himanshu P. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. // Materials Today: Proceedings. 2021. Volume 46. Part 20. p. 10922-10926.- (<https://www.sciencegate.app/document/10.1016/j.matpr.2021.01.948>)