

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
НЕКОММЕРЧЕСКОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ ИМЕНИ ГУМАРБЕКА
ДАУКЕЕВА»

ISSN 1999-9801

В Е С Т Н И К

АЛМАТИНСКОГО УНИВЕРСИТЕТА ЭНЕРГЕТИКИ И СВЯЗИ

Учрежден в июне 2008 года

Тематическая направленность: теплоэнергетика, электроэнергетика, радиотехника и связь,
космическая инженерия и технологии, информационные технологии, экология, обеспечение
жизнедеятельности, вопросы высшей технической школы

1 (52)

2021

Импакт-фактор - 0.105

Научно-технический журнал
Выходит 4 раза в год

Алматы

ВЕСТНИК АЛМАТИНСКОГО УНИВЕРСИТЕТА ЭНЕРГЕТИКИ И СВЯЗИ

СВИДЕТЕЛЬСТВО

О постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ KZ14VPY00024997

выдано

Министерством информации и общественного развития
Республики Казахстан

Подписной индекс – 74108

Бас редакторы – главный редактор

Стояк В.В.

К. т.н., профессор

Заместитель главного редактора
Ответственный секретарь

Жауыт Алгазы, доктор PhD
Шуебаева Д.А., магистр

Редакция алқасы – Редакционная коллегия

Сагинтаева С.С., д-р экон. наук, канд. физ.- мат. наук, академик МАИН, ректор НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Гита Ревалде, доктор PhD, член-корреспондент Академии наук Латвии, директор Национального Совета науки, Рига, Латвия

Главный редактор – Стояк В.В., канд. техн. наук, профессор, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Заместитель главного редактора – Жауыт А., доктор PhD, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Илиев И.К., д-р техн. наук, Русенский университет, Болгария

Белоев Кристо, д-р техн. наук, Русенский университет, Болгария

Галайко Дмитрий, доктор PhD, университет Сарбонны, Франция

Такая Инамори, доктор PhD, Университет Токио, Япония

Цветков В.Ю., д-р техн. наук, Белорусский государственный университет информатики и радиоэлектроники, Белоруссия.

Кузнецов А.А., д-р техн. наук, ФГБОУ ВА «Омский государственный университет путей сообщения», Российская Федерация.

Авезова Н.Р., д-р, техн. наук, Министерство инновационного развития Республики Узбекистан.

Мунц В.А., д-р техн. наук ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»,
Российская Федерация.

Мустафин М.А., д-р техн. наук, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Обозов А.Д., д-р техн. наук, Кыргызский государственный технический университет им. И. Раззакова, Кыргызская Республика

Орумбаев Р.К., д-р техн. наук, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Потехин В.В., канд. техн. наук доцент Санкт-Петербургского политехнического университета Петра Великого, г. Санкт-Петербург, Российская Федерация

Франческо Сандоро, доктор PhD, Университет Альдо Моро, Бари, Италия

Туманбаева К.Х., канд. техн. наук, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Мутуле Анна, доктор PhD, Рижский Технический Университет, Латвия

Махмутов С.К., канд. истор. наук, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Алипбаев К.А., доктор PhD, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

Кабдушев Б.Ж., канд. истор. наук, НАО «Алматинский Университет Энергетики и Связи имени Гумарбека Даукеева»

За достоверность материалов ответственность несут авторы.

При использовании материалов журнала ссылка на «Вестник АУЭС» обязательна.



АВТОМАТИКА, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И КИБЕРБЕЗОПАСНОСТЬ

УДК 004.93`12

https://doi.org/10.51775/1999-9801_2021_52_1_98

МҰНАЙ КЕН ОРЫНДАРЫН ИГЕРУДІ ОҢТАЙЛАНДЫРУ ҮШІН ДЕРЕКТЕРДІ ЖІКТЕУ

Н.М. Мейрамбек*, Е.Ғ. Кенжебек

әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
E-mail: nazerke.meyrambek@mail.ru

Аңдатпа. Берілген мақалада мұнай кен орындарын игеруді оңтайландыру үшін нейрондық желілердегі деректерді машиналық оқытуды жіктеу әдістері қарастырылады. Машиналық оқытудың әдістері ретінде шешім ағашы (Decision tree) және кездейсоқ орман әдісі (Random forest) алынады. Синтетикалық деректер жиынтығы Бакли-Леверетт математикалық моделін қолдану арқылы алынған. Бакли-Леверетт математикалық моделі жалпы мұнай өндіру мәселелеріндегі қанықтылық үлестірімін анықтау үшін қолданылады. Жіктеу алгоритмдерін тестілеу барысында деректер жиынтығының жалпы саны 134480 болды. Мұнай өндіру мәселесінің параметрлері ретінде әртүрлі комбинациялар таңдалды, онда кеуектілік, мұнай фазасының тұтқырлығы және абсолютті өткізгіштігі, мұнай беру коэффициентінің мәні алынды. Машиналық оқыту үшін Python бағдарламалау тілі пайдаланылды. Зерттеу нәтижелері Random forest әдісі Decision tree әдісіне қарағанда жоғары оқу нәтижелеріне ие екенін көрсетті.

Түйін сөздер: Машиналық оқыту, жіктелу, кездейсоқ орман әдісі, шешім ағашы әдісі, абсолютті өткізгіштік.

Кіріспе. Соңғы онжылдықта мұнай-газ саласында бұрғылауды басқару, жоғары ажыратымдылықты каротаж, телеметрия, барлау және пайдалану кезеңінде әртүрлі мәліметтер жинау және тағы басқалар сияқты процестерді компьютерлендіру жүргізілді. Бұл мақалада мұнай өнеркәсібіне арналған жаңа аналитикалық технологиялар қарастырылады[1].

Ұжымдық деректер модельдерін құру, талдау, процестерді модельдеу әдістерін өзгертуге мүмкіндік береді және тұтастай алғанда саланы оңтайландыруға ықпал етеді. Кен орындарын іздеудегі көптеген жетістіктер геология, петрофизика және геофизиканың үйлесімінен болды[2].

Машиналық оқыту әдістерін қолдана отырып, мұнай өндірісін жақсартуға байланысты көптеген жұмыстар бар. Осындай мақалалардың бірінде нейрондық желілердегі жіктеу бойынша мұнай және табиғи газ саласында тірек векторлық машиналарды(SVM) қолдануға шолу жасалған. Жіктелген белгілерді таңдау алгоритмдері: белгілерді саралау мен белгілер жиынтығы таңдалды[3]. Біріншіден, бұл зерттеуде ішкі жиынтықты таңдау мүмкін белгілердің кеңістігін тауларға өрмелеу, жан-жақты және генетикалық алгоритм және т.б. арқылы іздейді, оларды орамаларға, сүзгілерге бөлуге және ендіруге болады. Мұнай кен орындарын игеруде табиғи газды барлау және өндіру кезеңдерінде деректер көлемін екі санатқа бөлген. Сонымен қатар зерттеуде машиналық оқытуды жіктеудің 5 әдісі пайдаланылған: шешім ағашы (DT), екі түрі бар жасанды нейрондық желілер (ANN), тірек векторларының әдісі (SVM), байесиялық желілер (BS). Нәтижесінде тірек векторларының әдісі жақсы нәтижелер көрсетеді[4].

Келесі мақала авторлары машиналық оқыту әдісімен мұнай ұңғымаларының кенжар маңы аймақтарын өңдеу тиімділігін бағалауды қарастырған. Зерттеуде жіктелудің шешім ағашы әдісі пайдаланылған, 8 маңызды параметр саны үшін және 10000 операция жасалынады. Осылайша кен орнын игеруді басқаруды жүзеге асыру мақсатында шешім ағашынан тұратын алгоритм әзірленеді. Зерттеуде нейрондық желінің кіріс қабаты ретінде 7 параметрдің 2000 жиынтығы қолданылды. Мұнай кен орындарына болжам жүргізу барысында сызықты емес SVM алгоритмі мұнай ұңғымалары

жүйесін талдауда және тиімділікті болжауда сызықтық алгоритмге қарағанда жақсы жұмыс істейтіні дәлелденеді[5].

Машиналық оқыту әдістері мұнай-газ өнеркәсібінде кеңінен таралды және көптеген мәселелерді шешуде қолданылады. Алайда, осы мақаланың басты бағыты болып табылатын жергілікті жоспарлаудың маңызды мәселесіне жүгінетіндер аз. Бұл мақалада машиналық оқытуға негізделген далалық жоспарлау құрылымы ұсынылған және мұнай ұңғымаларының орналасуын жоспарлау мәселесі нақты қарастырылған. Бақыланбайтын оқыту деректердің сипаттамаларын түсіну үшін қолданылады, содан кейін ұңғымалардың өнімділігін болжау құралын жасау үшін маргиналды мұнай кен орнындағы мәліметтер негізінде дайындалған регрессиялық модель жасалады [6].

Бұл жұмыстың мақсаты мұнай кен орындарын игеруде нақты мәліметтерді жіктеуде машиналық оқытудың шешім ағашы және кездейсоқ орман әдістерін қолдану болып табылады. Екі әдіс үшін де болжау нәтижелері мен дәлдігін салыстыру.

Әдістер. Бұл бөлімде жұмыс процесі және машиналық оқытудың қолданылатын әдістері сипатталған. Сондай-ақ, мәліметтер жиынтығын құру, оқыту және тестілеу қарастырылады.

Бұл зерттеудің мақсаты мұнай кен орындарын игеру үшін машиналық оқыту әдістерін қолдану болып табылады. Синтетикалық мәліметтер жиынтығы гидродинамиканы есептеу және мұнай өндіру есептеріндегі қанықтылықтың таралуын анықтау үшін қолданылатын Бакли-Левверетт математикалық моделінің көмегімен алынды[7].

Зерттеуде Машиналық оқыту моделінің кіріс параметрлері ретінде 4 параметр алынды, ал резервуардағы мұнай шығару коэффициенті Шығыс параметрі ретінде алынды. Алынған мәліметтерде кіріс параметрлері ретінде мұнай өндіру міндеті параметрлерінің әртүрлі комбинациясы алынды: кеуектілік, мұнай фазасының тұтқырлығы және абсолютті өткізгіштігі. Шығу параметрі ретінде пласттың мұнай беру коэффициентінің мәні таңдалды.

Кіріс параметрлері:

- Кеуектілік – 41;
- Мұнай фазасының тұтқырлығы – 41;
- Өткізгіштігі – 2.

Осылайша, бұл жұмыста іріктеу жұптарының саны 3 362. Әрбір іске қосу кезінде пласттың мұнай беру коэффициентінің мәні есептеледі. Мұнайдың пайда болу коэффициентінің динамикасын зерттеу үшін үлгінің әр жұбы үшін аралық уақыттық итерация кезінде 40 мән алынды. Демек, мәліметтер жиынтығының жалпы саны 134 480.

Машиналық оқыту әдістері. Шешім ағашы (жіктеу ағашы немесе регрессия ағашы деп те аталады) - бұл машиналық оқыту, мәліметтерді талдау және статистикада қолданылатын шешімдерді қолдау құралы[8]. Ағаш құрылымы «жапырақтар» және «бұтақтар» арқылы ұсынылған. Шешім ағашының шеттері («бұтақтары») мақсат функциясы тәуелді атрибуттарды, «жапырақтары» мақсат функциясының мәндерін, ал басқа түйіндерде істер ажыратылатын атрибуттарын қамтиды. Жаңа жағдайды жіктеу үшін ағаштан жапыраққа түсіп, сәйкес мәнді қайтару керек. Осындай шешімді ағаштар деректерді өндіруде кеңінен қолданылады. Мақсат – бірнеше кіріс айнымалылар негізінде мақсатты айнымалының мәнін болжайтын модель құру[9].

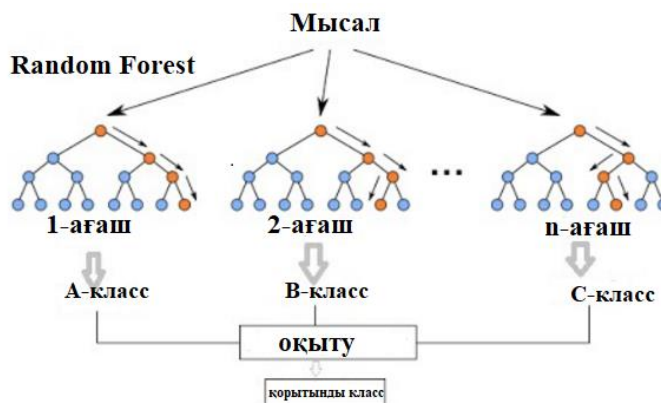
Нақты деректерді талдауда және жіктеуде оң дәлелденген белгілі әдістердің бірі – шешім қабылдау ағаштарының әдісі – машинаны оқытудың тиімді және ең бастысы интуитивті әдісі. Әдетте, шешім ағаштарын құру алгоритмдері жоғарыдан төменге жұмыс істейді. Сонымен қатар, әр кезеңде бақылаулардың бастапқы жиынтығын екі ішкі жиынға (сыныпқа) бөлетін ең таңдаулы фактор таңдалады[10].

Метрика ықтималдық сомасын білдіреді i индексі бар элементтің p_i , $\sum_{k \neq i} p_k = 1 - p_i$ ықтималдығына көбейтілген, берілген элементті таңдау қатесі. Егер барлық элементтер болса, метрика нөлге тең болады олар дұрыс таңдалмады және басқа біреудің санатына кірді. J кластарына түсетін көптеген элементтер үшін ластану мөлшерін есептеу үшін біз $i \in \{1, 2, \dots, J\}$ қоямыз және p_i i класқа түсетін элементтердің үлесі болсын. Келесі (1) өлшемді аламыз:

$$I_g(p) = \sum_{i=1}^J p_i \sum_{k \neq i} p_k = \sum_{i=1}^J p_i (1 - p_i) = \sum_{i=1}^J (p_i - p_i^2) = \sum_{i=1}^J p_i - \sum_{i=1}^J p_i^2 = 1 - \sum_{i=1}^J p_i^2. \quad (1)$$

Метрикананың ең жоғарғы мәні сәйкесінше белгінің ең таңдаулы екенін білдіреді. Әрі қарай, белгілі бір терминалдық шарттар орындалғанға дейін еншілес филиалдар үшін бірдей процесс жалғасады. Шешім ағаштарын құру процесі түйіндерде шешуші ережелерді қолдана отырып, оқу жиынтығын ішкі жиындарға жүйелі, рекурсивті бөлу болып табылады. Бөлу процесі барлық бұтақтардың соңындағы барлық түйіндер жапырақтары жарияланғанға дейін жалғасады. Түйіннің паракпен жариялануы табиғи түрде немесе пайдаланушы белгілеген тоқтау жағдайына жеткенде пайда болуы мүмкін [11].

Келесі пайдаланылған машиналық оқыту әдісі – кездейсоқ орман әдісі. Кездейсоқ орман – бұл шешімді ағаштардың ансамблі болып табылатын және шешім ағаштары бойынша пакетке салу әдісін жүзеге асыратын машиналық оқыту алгоритмі. Соңғысын оқытқан кезде, әр бөлім үшін ерекшеліктердің жаттығу үлгісі кейбір кездейсоқ ерекшеліктерден құралады. Кездейсоқ ормандарға арналған оқыту алгоритмі ағаштарды үйренушілерге сөмкелерді қалыптастырудың жалпы әдісін қолданады. Бір шешім ағашы бүкіл жаттығу жиынында жалғыз оқытылады. Кездейсоқ орманда N шешім ағаштары әрқайсысы бастапқы деректер жиынтығын бастапқы жүктеу арқылы алынған бастапқы оқыту жиынтығының ішкі жиынында, яғни ауыстырумен кездейсоқ іріктеу арқылы оқытылады [12]. Сонымен қатар, су объектілері бастапқы объектілер жиынтығының кездейсоқ жиынтығы ретінде ағаштан ағашқа дейін өзгеруі мүмкін. Әдетте, егер m – бастапқы деректер жиынтығындағы кіріс нысандарының саны болса, әр шешім ағашын оқыту үшін кездейсоқ алынған $[m$ квадрат түбірі] кіріс нысандарының жиынтығы қолданылады [13].



1 – сурет. Random Forest машиналық оқыту әдісінің жұмыс істеу технологиясы

Орманның кездейсоқ жіктеуіштерінің тиімді болуының басты себептерінің бірі – ормандағы әр ағаштың орманның басқа ағаштарымен көп дәрежеде байланысы жоқ және бұл жағдай әрбір жеке ағашты жаттықтыру үшін мәліметтерден кездейсоқ іріктеу (орау) арқылы жүзеге асырылады [14]. Нәтижесінде, кейбір ағаштар соңғы шығу класы үшін дұрыс емес болжамдар бере алады, бірақ біз ең көп ағаштар болжаған белгіні алғандықтан, біз өте жоғары дәлдікке ие боламыз, өйткені көптеген ағаштар дұрыс болжам жасай алады. Осылайша, бір-бірімен байланысты емес бірнеше шешуші ағаштарды қолдану біз жіберетін қателіктердің пайызын азайтуға көмектеседі, өйткені бір ағаштың жіберген қателіктеріне басқа ағаштар жол бермейді. Осылайша, біз тек бір дереккөзден емес, бірнеше көздерден болжамдар жасайтындықтан, біздің дәлдігіміз айтарлықтай артады [15].

Нәтижелер және талқылау. Машиналық оқытудың орындалу ортасы ретінде Python бағдарламалау тілі таңдалған болатын. Бұл бағдарламалау тілі машиналық оқытуға өте қолайлы, өйткені көптеген кітапханалары бар. Соның ішінде жұмыс істеу үшін пайдаланылған кітапханалар:

- Pandas кітапханасы мәліметтерді өңдеуге және талдауға арналған Python бағдарламалық кітапханасы пайдаланылды;
- Matplotlib – екі өлшемді (2D) графикамен деректерді визуализациялауға кітапхана;
- Numpy – бұл әр түрлі математикалық есептеулер мен көпөлшемді массивтермен жұмыс істеуге арналған кітапхана
- Scikit-learn – машиналық оқыту мәселелерін шешуге арналған ең көп таралған кітапхана.
- Graphviz – бұл әр түрлі түйіндер мен шеттердің көмегімен орындалатын графикалық объектілерді құру үшін қолданылатын ашық Python кітапханасы.

Мұнай кен орындарын игеруді оңтайландыру үшін деректерді жіктеу кезінде алдымен деректер қоры енгізілді (2-сурет). Нақты сыну жұмыстарынан алынған барлық мәліметтер базасын шартты түрде кіріс және шығыс мәліметтеріне бөлуге болады. Python бағдарламалық тілінде екі әдісті салыстырудың негізінде келесі мәндер шықты:

```
df.corr()
```

	Class	etta	Kviews	PORO	TIME	VISC_OIL
Class	1.000000	0.988491	NaN	-2.934715e-01	8.020694e-01	-4.041372e-01
etta	0.988491	1.000000	NaN	-2.944351e-01	8.121562e-01	-4.041561e-01
Kviews	NaN	NaN	NaN	NaN	NaN	NaN
PORO	-0.293472	-0.294435	NaN	1.000000e+00	-5.008123e-19	-5.862482e-18
TIME	0.802069	0.812156	NaN	-5.008123e-19	1.000000e+00	4.234532e-18
VISC_OIL	-0.404137	-0.404156	NaN	-5.862482e-18	4.234532e-18	1.000000e+00

```
X = df.iloc[:,1:]
X
```

	etta	Kviews	PORO	TIME	VISC_OIL
0	0.004099	2	0.130890	0	0.414938
1	0.028301	2	0.130890	200	0.414938
2	0.049026	2	0.130890	400	0.414938
3	0.068843	2	0.130890	600	0.414938
4	0.087731	2	0.130890	800	0.414938
...
67235	0.805976	2	0.107296	7000	0.115207
67236	0.818247	2	0.107296	7200	0.115207
67237	0.830421	2	0.107296	7400	0.115207
67238	0.842496	2	0.107296	7600	0.115207
67239	0.854474	2	0.107296	7800	0.115207

67240 rows x 5 columns

2-сурет. Мұнай кен орындарын игеруді оңтайландыру үшін деректерді жіктеудегі деректер қоры

Кездейсоқ орман алгоритмі - бұл жеке өсірілген шешім ағаштарының жиынтығы ғана емес. Random Forest (RF) алгоритмі бірнеше ағаштардан туындаған кездейсоқтықтан басқа, жеке ағаштарды екі түрлі тәсілмен салу кезіндегі кездейсоқтықты қамтиды, олардың ешқайсысы қарапайым шешімдер ағашы (DT) алгоритмінде жоқ. Біріншісі - ағаштың әр түйінінде ең жақсы бөлуді іздеу кезінде ескерілетін мүмкіндіктер саны: DT барлық мүмкіндіктерді қарастырса, RF олардың кездейсоқ ішкі жиынын қарастырады. Екіншіден, DT жаттығу жиынтығын қарастырған кезде, жалғыз RF ағашы тек оның жүктелген ішкі үлгісін қарастырады.

```
from sklearn.ensemble import RandomForestRegressor

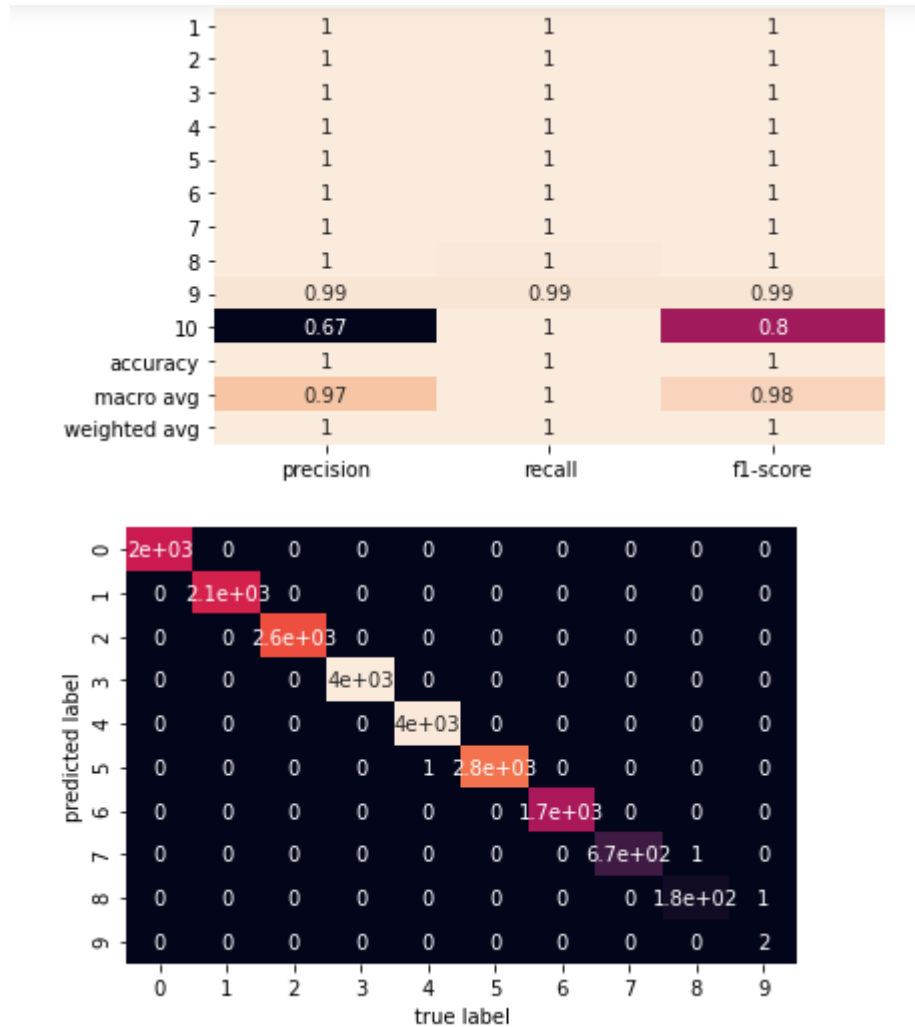
regressor = RandomForestRegressor(n_estimators=20, random_state=0)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
```

```
from sklearn import metrics
print("Accuracy:", metrics.accuracy_score(y_test, y_pred.round()))
```

Accuracy: 0.9998512790005949

3-сурет. Random Forest әдісінің берілген нәтижелері

Python бағдарламалау тілінде кездейсоқ орман алгоритмінің нәтижелері алынды(3-сурет). Берілген нәтижелер есебінен алынған мұнай кен орындарында мәліметтерді жіктеу корреляциялық матрицасы(4-сурет) көрсетілді.



4 – сурет. Random Forest әдісімен жасалынған корреляциялық матрицаларды салыстыру

Шешімдер ағаштарын құру үшін әр түйінде ең жақсы таңдауды анықтай алатын алгоритмдер қажет. Танымал алгоритмдердің бірі - Hunt алгоритмі. Бұл ашкөз модель, яғни әр кезеңде ең жақсы шешім қабылдайды, бірақ жаһандық оптимумды ескермейді. Бұл нені білдіреді? Әр қадамда алгоритм ең жақсы нәтижені таңдайды. Алайда, осы кезеңде ең жақсы нәтижені таңдау ағаштың жапырақ түйіні деп аталатын соңғы түйініне жеткенде оңтайлы шешімге әкелетін жолмен жүруге кепілдік бермейді. Шешім беретін ағаштар, әсіресе ағаш өте терең болған кезде, артық отыруға бейім. Бұл алдыңғы жорамалдарға сәйкес келетін оқиғалардың кішігірім үлгісіне әкелетін біз қарастыратын ерекшелік дәрежесіне байланысты. Бұл шағын үлгі қате тұжырымдарға әкелуі мүмкін. Python бағдарламалау тілінде шешімдер ағаштар алгоритмінің нәтижелері алынды(5-сурет).

```
clf = DecisionTreeClassifier()
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
```

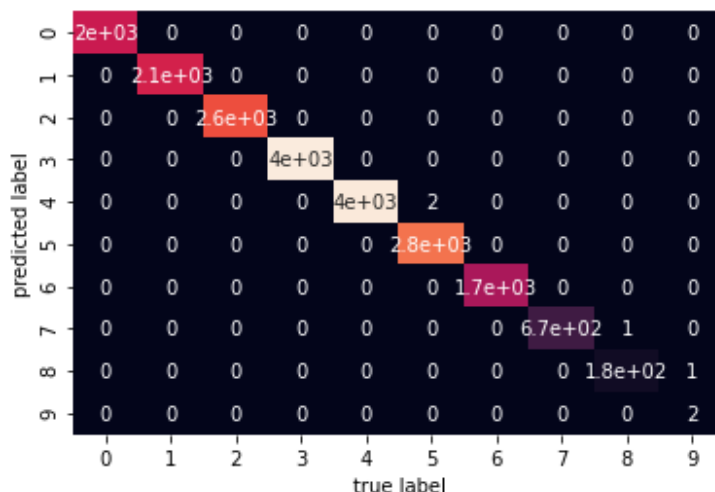
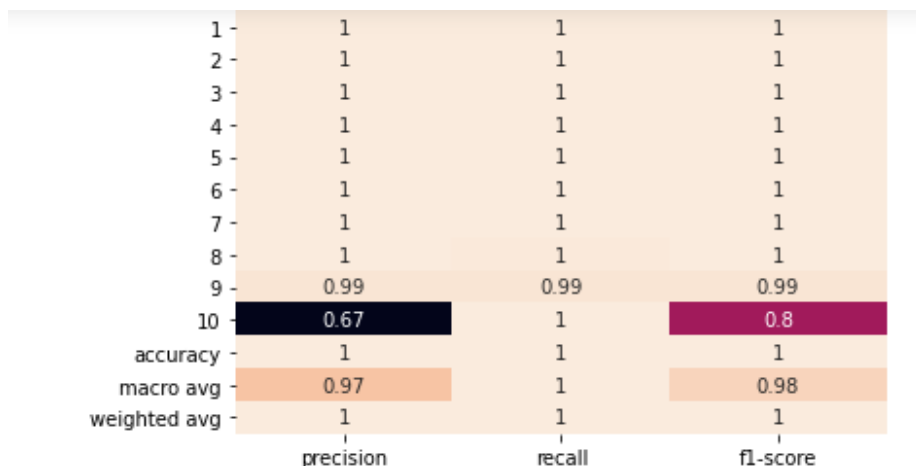


```
print("Accuracy:", metrics.accuracy_score(y_test, y_pred))
```

Accuracy: 0.9998017053341265

5-сурет. Decision tree әдісінің берілген нәтижелері

Келесі кескінде соңғы әдіс нәтижелер есебінен алынған мұнай кен орындарында мәліметтерді классификациялау корреляциялық матрицасы көрсетілді(6-сурет).



6 – сурет. Decision tree әдісімен жасалынған корреляциялық матрицаларды салыстыру

Ақ түс айнымалылардың корреляцияға жатпайтынын көрсетеді, ал қара қызылға қарай байланыс неғұрлым көбірек болса, кері корреляция соғұрлым көп болады.

Бұл жерде корреляциялық талдаудың міндеттері:

- а) екі немесе одан да көп құбылыстардың қосылу дәрежесін.
- ә) құбылыстар арасындағы байланыс дәрежесін өлшеуге негізделген тиімді атрибутқа едәуір әсер ететін факторларды таңдау. Осы тұрғыдан маңызды факторлар регрессиялық талдау кезінде одан әрі қолданылады.
- в) Белгісіз себеп-салдарлық байланыстарды анықтау.

Қарым-қатынастың көріну формалары өте алуан түрлі. Функционалды (толық) және корреляциялық (толық емес) байланыстар олардың кең таралған түрлері ретінде ажыратылады.

Синтетикалық деректер жиынтығы оқу және тестілік мәліметтер жиынтығына бөлінді. Оқыту үшін жалпы деректердің 2353 жиынтығы (70%), ал тестілеу үшін қалған 1009 жұп (30%) пайдаланылды. Жоғарыда айтылғандай, үлгі жұптарының жалпы саны 10086 модельді құрайды. Іріктеменің әрбір жұбы мұнай беру коэффициентінің 40 мәнінен тұрады. Нәтижесінде бізде көптеген тест жұптары бар, бірақ бұл жұмыста нәтижелер кейбір тест жұптары үшін көрсетіледі. 1 – кестеде

көрсетілген нәтижелер Python бағдарламалау тілінде оқытылған 30% тестіленген және 70% оқытылған нәтижелер көрсетілген.

1-кесте

Машиналық оқыту классификатор түрі	Макроорта бағалау (macro – averages)	Дәлдік (accuracy)
Decision tree	0,97	0,999801
Random forest	0,98	0,999851

Қорытынды

Бұл мақала мұнай кен орындарын игеруді оңтайландыру үшін деректерді жіктеу кезінде машиналық оқыту әдістерін қолдануға арналды. Бұл зерттеуде машиналық оқытудың жіктелу әдістері ретінде шешім ағашы мен кездейсоқ орман әдістері қолданылды. Нәтижесінде, кездейсоқ ормандар - бұл модельдеудің күшті әдісі және бір шешім ағашына қарағанда әлдеқайда берік. Олар көптеген шешімдерді біріктіреді, бұл артық сәйкестікті, сондай-ақ қателіктерді шектейді, сондықтан пайдалы нәтижелер береді. Зерттеу мақала нәтижесі бойынша кездейсоқ ормандар әдісінің тиімді екенін көрсетті.

ӘДЕБИЕТТЕР ТІЗІМІ

- [1] Применение машин опорных векторов на этапе разведки нефти и природного газа: кандидат системотехники, аспирантура в области промышленной и системной инженерии, Национальный инженерный университет Av. Túpac Amaru 2010 - Римак / Лима 25 - Перу.
- [2] Курганов Д.В. Расчет эффекта от перевода добывающей нефтяной скважины в нагнетательный фонд в рамках управления разработкой нефтяным месторождением // Управление большими системами. 2019. - № 81 - С. 147–167.
- [3] Матс, Г. Андерсен «Оптимизация добычи пласта с использованием генетических алгоритмов и искусственных нейронных сетей»; M.Sc. Диссертация 2009 г., Норвежский университет науки и технологий.
- [4] Моделирование периодической работы добывающей скважины, эксплуатирующей послойно неоднородный по проницаемости пласт залежи высоковязкой нефти / И.В. Владимиров, О.Н. Пичугин, П.Н. Соляной, А. Хисаева // Вестник ЦКР Роснедра, 2015. - №1. - С. 23-29.
- [5] Ф. Анифоуз и А. Абдулрахим, «Слияние функциональных сетей и нечеткой логики типа 2 для характеристика нефтегазовых коллекторов» в Материалы Международной конференции 2010 г. Электроника и информационная инженерия (ICEIE 2010), Киото, Япония, 1-3 августа, IEEE Xplore Vol. 2, 2010, стр. 349-353.
- [6] Бахши Х., Дехгани А., Джафарипанах С. Использование генетического алгоритма на основе уравнения Риделя для прогнозирования давления пара органических соединений. Международный журнал инженерии. 2018 г. №6: 863-9.
- [7] Повышение эффективности обнаружения выбросов для структур с низкой плотностью / Дж. Тан, З. Чен, А.В. Фу, Д.В. Чунг / Ин: Чен М.С., Ю П.С., Лю Б. (ред.) Достижения в области открытия знаний и данных добыча полезных ископаемых. Конспект лекций по информатике ПАКДД. - Спрингер. - 2002. - Т. 2336. - С. 535–548.
- [8] Валеев С.А., Дулкарнаев М.Р., Бриллиант Л.С. Методологические основы планирования и организации систем интенсивного заводнения // Экспозиция Нефть Газ. 2016. № 3 (49).
- [9] Подольский А.К. Применение методов искусственного интеллекта в нефтегазовой отрасли // Современная наука. - 2016. - №3.

[10] Трент Дж. Аналитические фирмы изучают рынок нефти и газа. Журнал нефтяных технологий. Октябрь 2016. 36–38.

[11] Рябец Д.А., Бриллиант Л.С., Завьялов А.С. Управление добычей на основе нейросетевой оптимизации работы скважин на объекте БС8 Западно-Малобалыкского месторождения // Нефтегаз.РУ. 2019. №6.

[12] Спесивцев П., Синьков К., Софронов И., Зимица А., Умнов А., Яруллин Р., Д. Ветров. Прогнозная модель забойного давления на основе машинного обучения. Журнал нефтегазовой науки и техники, 166: 825–841, 2018 г.

[13] Иванов Д.В., Чижов А.П. Методы интенсификации добычи нефти. Сборник научных трудов «Нефтегазовые технологии и новые материалы. Проблемы и решения» [Сборник научных трудов «Нефтегазовые технологии и новые материалы. Проблемы и решения»]. Уфа, 2014, вып.3 (8), с. 136-139. (на русск. яз.).

[14] Боровиков В.П. Нейронные сети. STATISTICA Neural Networks: методология и технологии современного анализа данных. 2-е изд., Москва: Горячая линия - Телеком, 2008.

[15] Абишев А.А., Воробьев А.Е., Тчаро Х. Перспективы цифровизации нефтяной отрасли Казахстана // Вестник АУНГ (Казахстан) N 1 (45). 2018. С. 37-46.

REFERENCES

[1] Application of Support Vector Machines in the Oil and Natural Gas Exploration Phase: PhD in Systems Engineering, Graduate School of Industrial and Systems Engineering, National University of Engineering Av. Túpac Amaru 2010 - Rimac / Lima 25 - Peru.

[2] Kurganov D.V. Calculation of the effect of transferring a producing oil well to an injection fund within the framework of oil field development management // Management of large systems. 2019. - No. 81 - pp. 147–167.

[3] Mats, G. Andersen “Optimization of Reservoir Production Using Genetic Algorithms and Artificial Neural Networks”; M.Sc. Dissertation 2009, Norwegian University of Science and Technology.

[4] Modeling of periodic operation of a producing well operating a layer-by-layer heterogeneous reservoir of high-viscosity oil / I.V. Vladimirov, O. N. Pichugin, P.N. Solyanoy, A. Khisaeva // Bulletin of the Central Committee for Regional Development of Rosnedra, 2015. - No. 1. - S. 23-29.

[5] F. Anifouz and A. Abdulrahim, "Fusion of Functional Networks and Type 2 Fuzzy Logic to Characterize Oil and Gas Reservoirs" in Proceedings of the 2010 International Conference on Electronics and Information Engineering (ICEIE 2010), Kyoto, Japan, August 1-3, IEEEExplore Vol. 2, 2010, pp. 349-353.

[6] Bakhshi H., Dehgani A., Jafaripana S. Using a genetic algorithm based on the Riedel equation to predict the vapor pressure of organic compounds. International Engineering Journal. 2018 No. 6: 863-9.

[7] Improving the efficiency of detecting outliers for structures with low density / J. Tan, Z. Chen, A.V. Fu, D.V. Chung / Ying: Chen M.S., Yu P.S., Liu B. (ed.) Advances in the discovery of knowledge and mining data. Lecture notes on PACDD informatics. - Springer. - 2002. - T. 2336. - S. 535–548.

[8] Valeev S.A., Dulkarnaev M.R., Brilliant L.S. Methodological bases of planning and organization of intensive waterflooding systems // Exposition Oil Gas. 2016. No. 3 (49).

[9] Podolsky A.K. Application of artificial intelligence methods in the oil and gas industry // Modern Science. - 2016. - No. 3.

[10] Trent J. Analytical firms study the oil and gas market. Journal of Petroleum Technology. October 2016.36–38.

[11] Ryabets D.A., Brilliant L.S., Zavyalov A.S. Production management based on neural network optimization of wells at the BS8 facility of the West Malobalykskoye field // Neftegaz.RU. 2019. No. 6.

[12] Spesivtsev P., Sinkov K., Sofronov I., Zimina A., Umnov A., Yarullin R., D. Vetrov. Downhole pressure predictive model based on machine learning. Journal of Oil and Gas Science and Technology, 166: 825–841, 2018

[13] Ivanov D.V., Chizhov A.P. Oil production intensification methods. Collection of scientific papers “Oil and gas technologies and new materials. Problems and solutions “[Collection of scientific papers" Oil and gas technologies and new materials. Problems and solutions ”]. Ufa, 2014, issue 3 (8), p. 136-139. (in Russian).

[14] Borovikov V.P. Neural networks. STATISTICA Neural Networks: Methodology and Technologies of Modern Data Analysis. 2nd ed., Moscow: Hotline - Telecom, 2008.

[15] Abishev AA, Vorobiev AE, Tcharo H. Prospects for digitalization of the oil industry in Kazakhstan // Bulletin AUNG (Kazakhstan) N 1 (45). 2018.S. 37-46.

DATA CLASSIFICATION TO OPTIMIZE OIL FIELD DEVELOPMENT

N.M. Meyrambek*, Y.G. Kenzhebek

Al-Farabi Kazakh National University, Almaty, Kazakhstan

E-mail: nazerke.meyrambek@mail.ru

Abstract. *Methods for classifying machine learning data in neural networks to optimize the development of oil fields are discussed in this article. We took the Decision tree and the Random forest method as the machine learning methods. The synthetic dataset was obtained using the Buckley-Leverett mathematical model. The Buckley-Leverett mathematical model is used to determine the saturation distribution in general oil production problems. In the course of testing the classification algorithms, the total number of data sets was 134,480. Various combinations were chosen as parameters of the oil production problem, in which the porosity, viscosity and absolute permeability of the oil phase, the value of the oil recovery factor were obtained. For machine learning, the Python programming language was used. The research results showed that the Random forest method has higher learning outcomes than the Decision tree method.*

Key words. *Machine learning, classification, random forest method, decision tree method, absolute conductivity.*

КЛАССИФИКАЦИЯ ДАННЫХ ДЛЯ ОПТИМИЗАЦИИ РАЗРАБОТКИ НЕФТЯНЫХ МЕСТОРОЖДЕНИЙ

Н.М. Мейрамбек*, Е.Г. Кенжебек

Казахский национальный университет имени аль-Фараби, Алматы, Казахстан

E-mail: nazerke.meyrambek@mail.ru

Аннотация. *В данной статье рассматриваются методы классификации машинного обучения данных в нейронных сетях для оптимизации разработки нефтяных месторождений. В качестве методов машинного обучения берется дерево решений (Decision tree) и метод случайных лесов (Random forest). Синтетический набор данных получен с помощью математической модели Бакли-Левверетта. Математическая модель Бакли-Левверетта используется для определения распределения насыщения в общих задачах добычи нефти. В ходе тестирования алгоритмов классификации общее количество наборов данных составило 134480. В качестве параметров проблемы нефтедобычи выбраны различные комбинации, в которых получены пористость, вязкость и абсолютная проницаемость нефтяной фазы, значение коэффициента нефтеотдачи. Для машинного обучения использовался язык программирования Python. Результаты исследования показали, что метод Random forest имеет более высокие результаты обучения, чем метод Decision tree.*

Ключевые слова: *машинное обучение, классификация, метод случайного леса, метод дерева решений, абсолютная проницаемость.*

Басылымның шығыс деректері

Мерзімді баспасөз басылымының атауы	«Алматы энергетика және байланыс университетінің Хабаршысы» ғылыми-техникалық журналы
Мерзімді баспасөз басылымының меншік иесі	«Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» коммерциялық емес акционерлік қоғамы
Бас редактор	Профессор, т.ғ.к., В.В. Стояк
Қайта есепке қою туралы куәліктің нөмірі мен күні және берген органның атауы	№ KZ14VPY00024997, күні 17.07.2020, Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігі
Мерзімділігі	Жылына 4 рет (тоқсан сайын)
Мерзімді баспасөз басылымының реттік нөмірі және жарыққа шыққан күні	Жалпы нөмір 53, 2-басылым, 2021 жылғы 30 маусым
Басылым индексі	74108
Басылым таралымы	200 дана
Баға	Келісілген
Баспахана атауы, оның мекен-жайы	«Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» КЕАҚ баспаханасы, Байтұрсынұлы көшесі, 126/1 үй, А120 каб.
Редакцияның мекен-жайы	050013, Алматы қ., «Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» КЕАҚ, Байтұрсынұлы к-сі, 126/1 үй, каб. А 224, тел.: 8 (727) 292 58 48, 708 880 77 99, e-mail: vestnik@aes.kz

Выходные данные

Название периодического печатного издания	Научно-технический журнал «Вестник Алматинского университета энергетике и связи»
Собственник периодического печатного издания	Некоммерческое акционерное общество «Алматинский университет энергетике и связи имени Гумарбека Даукеева»
Главный редактор	Профессор, к.т.н., Стояк В.В.
Номер и дата свидетельства о постановке на учет и наименование издавшего органа	№ KZ14VPY00024997 от 17.07.2020 Министерство информации и общественного развития Республики Казахстан
Периодичность	4 раза в год (ежеквартально)
Порядковый номер и дата выхода в свет периодического печатного издания	Валовый номер 53, выпуск 2, 30 июня 2021
Подписной индекс	74108
Тираж выпуска	200 экз.
Цена	Договорная
Наименование типографии, ее адрес	Типография НАО «Алматинский университет энергетике и связи имени Гумарбека Даукеева», ул. Байтұрсынұлы, дом 126/1, каб. А 120
Адрес редакции	050013, г. Алматы, НАО «Алматинский университет энергетике и связи имени Гумарбека Даукеева», ул. Байтұрсынұлы, дом 126/1, каб. А 224, тел.: 8 (727) 292 58 48, 708 880 77 99, e-mail: vestnik@aes.kz

Issue output

Name of the periodical printed publication	Scientific and technical journal "Bulletin of the Almaty University of Power Engineering and Telecommunications"
Owner of the periodical printed publication	Non-profit joint-stock company "Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeev"
Chief Editor	Professor, candidate of technical sciences Stoyak V.V.
Number and date of the registration certificate and the name of the issuing authority	№ KZ14VPY00024997 from 17.07.2020 Ministry of Information and Social Development of the Republic of Kazakhstan
Periodicity	4 times a year (quarterly)
Serial number and date of publication of a periodical printed publication	Number 53, edition 2, June 30, 2021
Subscription index	74108
Circulation of the issue	200 copies
Price	Negotiable
The name of the printing house, its address	Printing house of Non-profit joint-stock company "Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeev", 126/1 Baitursynuly str., office A 120
Editorial office address	050013, Non-profit joint-stock company "Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeev", A 224, tel.: 8 (727) 292 58 48, 708 880 77 99, e-mail: vestnik@aes.kz