

**ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ**

**«Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ» КЕАҚ**

**Студенттер мен жас ғалымдардың  
«GYLYM JÁNE BILIM - 2024»  
XIX Халықаралық ғылыми конференциясының  
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ  
XIX Международной научной конференции  
студентов и молодых ученых  
«GYLYM JÁNE BILIM - 2024»**

**PROCEEDINGS  
of the XIX International Scientific Conference  
for students and young scholars  
«GYLYM JÁNE BILIM - 2024»**

**2024  
Астана**

**УДК 001**

**ББК 72**

**G99**

**«ǴYLYM JÁNE BILIM – 2024» студенттер мен жас ғалымдардың XIX Халықаралық ғылыми конференциясы = XIX Международная научная конференция студентов и молодых ученых «ǴYLYM JÁNE BILIM – 2024» = The XIX International Scientific Conference for students and young scholars «ǴYLYM JÁNE BILIM – 2024». – Астана: – 7478 б. - қазақша, орысша, ағылшынша.**

**ISBN 978-601-7697-07-5**

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

**УДК 001**

**ББК 72**

**G99**

**ISBN 978-601-7697-07-5**

**©Л.Н. Гумилев атындағы Еуразия  
ұлттық университеті, 2024**

Осы зерттеулердің нәтижесінде компьютерлік көру технологиясын қолдана отырып, тұрмыстық қатты қалдықтарды жіктеу және сұрыптау жүйесі әзірленді. Қалдықтарды басқаруда компьютерлік көруді пайдалану сұрыптау процестерін автоматтандыруға және қалдықтарды басқару тиімділігін арттыруға мүмкіндік беретін перспективалы бағыт болып табылады.

Зерттеудің негізгі нәтижелеріне мыналар жатады:

Терең оқытуды пайдалана отырып, кескінді талдау негізінде қалдықтарды жіктеу алгоритмін құру. Бұл алгоритм пластмасса, қағаз, металл және басқалар сияқты қалдықтардың түрлерін дәл анықтауға мүмкіндік береді, бұл оларды кейіннен сұрыптаудың негізгі қадамы болып табылады.

Жіктеу нәтижелері бойынша қалдықтарды сұрыптау жүйесін құру. Жүйе қалдықтарды кейін өңдеу немесе кәдеге жарату үшін автоматты түрде сәйкес контейнерлерге таратады.

Эксперимент жүргізу және нақты деректер бойынша әзірленген жүйені сынау. Алынған нәтижелер жоғары классификация дәлдігін және ұсынылған технологияны қолдану арқылы қалдықтарды сұрыптау тиімділігін көрсетеді.

Тұрмыстық қатты қалдықтарды жіктеудің және сұрыптаудың ұсынылып отырған жүйесі қалдықтарды өңдеудің тиімділігі мен экологиялық тұрақтылығын арттыруға көмектесетін қалдықтарды басқару процестерін автоматтандыруға бағытталған маңызды қадам болып табылады. Әрі қарайғы зерттеулер жүйенің функционалдығын кеңейтуге, сондай-ақ оны нақты жағдайларда кеңінен қолдану үшін өндірістік және қалалық инфрақұрылымдарға біріктіруге бағытталуы мүмкін.

#### **Қолданылған әдебиеттер тізімі**

1. Рейнхард Клетте. «Компьютерное зрение. Теория и алгоритмы» / пер. с англ. А. А. Слинкин. – М.: ДМК Пресс, 2019. – 506 с.
2. Николенко С., Кадури А., Архангельская Е. Н63 Глубокое обучение. — СПб.: Питер, 2018. — 480 с.
3. Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Изд-во Казан. ун-та, 2018. – 121 с.

ӘОЖ 004.021

#### **ТАЛАПкерлерді жоо-ға қабылдауды болжаудың алгоритмдік үлгілеріне шолу**

**Умарова Зухра Мухитқызы**

**zukhra.umarova03@mail.ru**

Л.Н.Гумилев атындағы ЕҰУ Компьютерлік және программалық инженерия  
кафедрасының магистранты, Астана, Қазақстан

Ғылыми жетекші – Г.М. Баенова

Бұл мақаланың негізгі мақсаты – гранттар бойынша жоғары оқу орындарына түсетін студенттердің санын болжау үшін машиналық оқытудың тиімді алгоритмдерін қарастыру. Мақалада статистикалық деректерден алынған үлгілер мен статистикалық талдау әдістері қарастырылады. Жоо-ға түсуші талапкерлердің санына әсер ететін факторларды ескере отырып, алгоритмдердің үлгілері зерттелді.

Кілт сөздер: машиналық оқыту, Python, статикалық деректер, модель, алгоритм, талапкер.

#### **Кіріспе**

Ақпараттық технологиялардың дамуымен олардың әсер ету және қолдану аясы адам өмірінің барлық салаларында, оның ішінде оқу-тәрбие процесінде қазіргі қоғам өмірінің барлық деңгейлерін қамтыды. Заманауи нарық жағдайында жоғары оқу орындарының дамуы ел болашағының техникалық-экономикалық стандарттар бойынша көрсеткішіне айтарлықтай әсер етеді.

Көп жылдар бойы оқуға түсу кезінде студенттердің көпшілігінің таңдауы белгілі бір университеттерге түсетіні айғақ. Бұл студенттер санының теңгерімсіздігіне әкеліп соқтырғанымен, бұл тұрақты деректер аз таңдалымға ие болған бірнеше университеттердегі білім сапасын арттыруға қызмет етеді. Абитуриенттер, көбінесе, рейтинг негізінде жоғары оқу орындарына таңдау жасайды. Ең танымал рейтинг Британдық Times Higher Education (THE) – ең алдымен бес бағыт бойынша (оқыту, ғылыми-зерттеу, дәйексөздер, халықаралық өзара әрекеттесу, өндірістік қызметтен түскен табыс) университеттерді бағалауға негізделсе, тағы бір қолданыстағы рейтинг Quacquarelli Symonds (QS) ғалымдар мен жұмыс берушілер арасындағы университеттің беделін бағалау үшін сауалнама жүйесін пайдаланады, сонымен қатар, университеттің кадрлық деңгейіне қызығушылық танытады [1]. Сондай-ақ, оқушылардан алынған жауаптар бойынша таңдау жасауға “сапалы білім” - 33%, “жақсы бедел” - 30%, “білікті ұстаздар” - 25%, “университеттің үйге жақын болуы” - 21%, “мансаптық өсу” - 22%, және т.б. факторлар әсер етеді [2].

Өте күрделі үлгілерді анықтау үшін іске асырылуы мүмкін машиналық оқыту алгоритмдерінің кең ауқымы бар. Берат және басқа ғалымдардың зерттеуінше, дәлдік әрқашан қол жетімді деректер жинағының сапасына байланысты, ол деректер ғылымының ең жақсы тәжірибелерін орындау арқылы алдын ала өңделуі керек [3]. Әрине, әртүрлі сипаттағы мәселелер әртүрлі әдістерді қажет етеді: бір алгоритм барлық деректер жиынына қолдануға және жоғары дәлдіктегі корреляцияларды шығаруға жарамайды. Осы себепті абитуриенттердің таңдауы бойынша көзделген қанағаттанарлық өнімділікке жеткенше көптеген алгоритмдер сыналып, бір-бірімен салыстырылды.

### **Зерттеу**

Адамдардың қандай да бір таңдау жасауында болжамдардың шынайылығын машиналық оқыту әдістері мен дәстүрлі әдістер арқылы салыстыру Мартинес-Гармендияның еңбегінде зерттелген [4]. Екі әдістің ең үлкен тұжырымдамалық айырмашылығы - бірінші әдіс гипотезалар стильдендірілген математикалық модель түрінде айқындалады да, кейін қайта тексеріледі; ал екінші әдіс болжау дәлдігін барынша арттыруға тырысады, сондықтан нақты теориялық негіздерді ескермей, математикалық конструкцияларда үлкен икемділікті қамтамасыз етеді.

Модельдер таңдалатын өнімнің даралығын және гетерогенділігін көрсету мақсатында бұл зерттеуде сыра санатындағы сатып алулар панель және дүкен сканерлері деректері біріктірілді. Дүкеннің он екі жылдық деректері (2001-2012) 47 базардағы желілік азық-түлік дүкендері мен дәріханалар үшін ұсынылған. Ол сонымен қатар екі BehaviorScan IRI нарығына (11 жаста, сол санаттар үшін) арналған панельді сатып алу деректерін және демографиялық деректерді ұсынады. Өнімнің сатып алу болжауының дұрыстығын тексеру үшін 2001 жылғы деректер 2002 жылға болжаушы ретінде қолданды. Тұтынушылардың таңдауын болжау классификациялық модель арқылы орындалды.

Сипаттамалық статистикалық нәтижелерде оқуға түсетін студенттердің арасында менеджмент дәрежесі бағдарламасы ең көп қызығушылық тудырғаны және оқу жоспарлары жаңа болмағандықтан, балық шаруашылығы бойынша білім беру бағдарламасы келген студенттер арасында ең аз қызығушылық тудырғаны анықталды. Кластерлік талдау нәтижелері бойынша 8 топ құрылды. Менеджмент, өнеркәсіптік инженерия және PGSD бағдарламаларынан тұратын 5-топ бір оқу бағдарламасына шаққандағы студенттердің орташа саны бойынша көбірек қызығушылық танытты. Жылжымалы ортаны пайдаланатын болжам да бұл топтың өсу үрдісі бар екенін дәлелдейді. XXX университетіне жаңа студенттерді қабылдауға деген қызығушылықты болжау үшін қолданылатын статистикалық әдіс бір айнымалы деректерді болжауда жоғары дәлдігі белгілі авторегрессивті біріктірілген жылжымалы орташа (ARIMA) болжау үлгісі болып табылады. Бұл зерттеуде XXX университетінің басынан соңғы қол жетімді деректерге дейін оқуға түскен жаңа студенттердің саны туралы деректер пайдаланылады [5].

Студенттердің жетістіктерін болжауда олардың алдыңғы 10-13 сыныптағы бағаларын, мектеп бітіру деңгейін, қабылдану емтиханының көрсеткіштерін негізге ала отырып 16

алгоритм есебінде түрлі машиналық оқытудың 4 әдісінде модель жасаған болатын. Weka бағдарламасында деректер қорын сараптаумен іске асырылған бұл модельдерің негізгі мақсаты студенттерді екі топқа бөлу - мықты және әлсіз.

Сондай-ақ, студенттердің оқудағы жетістіктері негізінде оқудан шығуын болжамдайтын зерттеу жұмысында бірнеше академиялық бағдарламалар мен бір университеттегі жаңа студенттердің қызығушылықтарының санының болжамына тәуелді болып табылады. Сан-Маркос Ұлттық Мэри Университетінің (FISI-UNMSM) жүйелік инженерия және информатика факультетіндегі студенттердің оқуды тастап кету деңгейін деректерді іздеу тәсілімен және интеллектуалды жүйені енгізу арқылы азайту тәсілін ұсынады [6]. Қай студенттердің оқуды тастап кету қаупі бар екенін болжайтын автономды оқыту алгоритмі (шешім ағаштары) Python-да және Weka тегін бағдарламалық жасақтамасында жасалған. Ол үшін 2004-2014 жылдар аралығында факультетке түскен студенттердің мәліметтері қарастырылды. Бұл шешім факультеттің қолжетімділігін және қанағаттану деңгейін арттырады; оқу процесінде 90,34% дәлдік пайызы және 95,91% дәлдік алынды, сондықтан деректерді іздеу моделі жарамды болып саналады.

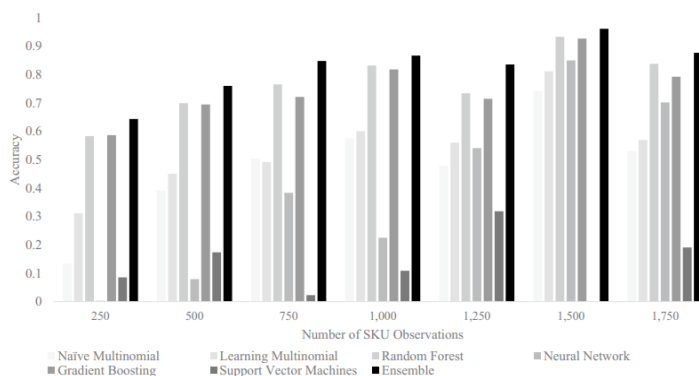
### Зерттеу нәтижелері

Модель параметрлерін интерпретациялаудың орнына болжау қабілетіне баса назар аударатын машиналық оқытудың модельдеуінен кейін әрбір модель үшін жай баптау параметрлері хабарланды. Жеті модель үшін баптау параметрлері 1-кестеде келтірілген.

1-кесте. Алгоритмдер бойынша параметрлерді баптау

Тәсіл	Параметр	Мәні	Минималд ы	Орташа	Максималд ы	Дисперсия
Naïve multinomial	NA	NA	0.00	0.22	0.89	0.06
Learning multinomial	$\lambda$	0.0001	0.00	0.37	0.89	0.05
Random forests	mtry B	68 50	0.00	0.62	1.00	0.04
Neural networks	H $\lambda$	7 0.3	0.00	0.05	0.85	0.03

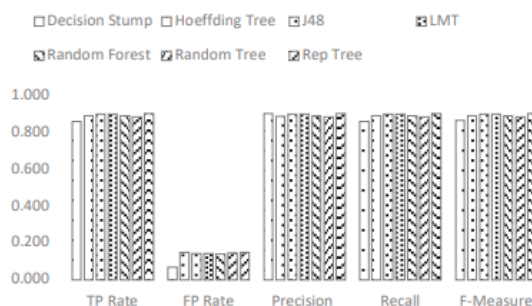
Градиентті күшейтуді қолдану нәтижесінде алынған дәлдік пен Карра салыстырмалы түрде жоғары, кездейсоқ орман үлгісі қол жеткізген мәндерден ұялшақ және эталондық аңғал көпмүшелік логит үлгісіне қарағанда жақсырақ. Сезімталдық тіпті кездейсоқ орманға қарағанда орта есеппен жоғары. Дәлдік салыстырмалы түрде сенімді деңгейлерді аз бақылаулардың өзінде көрсетеді (Сурет 1), бірақ әлі де көп бақылаулардан пайда көреді.



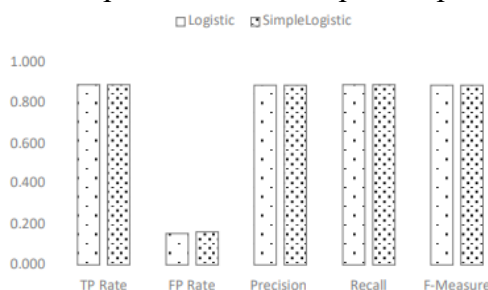
1-сурет. SKU бақылауларының саны бойынша әдіс дәлдігі

Бұл зерттеу қатысушылардың келесі SKU сатып алуларын болжау үшін машиналық оқыту үлгілерінің өнімділігін бағалайды. Аңғал көпмүшелік логит үлгісі сезімталдығы 0,00-ден 0,89-ға дейін және 0,37 дәлдігімен эталон ретінде қызмет етеді, бұл SKU бақылауларының жоғарылауымен айтарлықтай жақсарады. Оқытудың көпмүшелік логиті моделі 0,0001 оңтайлы ыдырау баптау параметрі ( $\lambda$ ) және сезімталдық пен дәлдіктің орташа жақсартуларымен аңғал модельден асып түседі.  $mtry=68$  және  $B=50$  баптау параметрлері бар кездейсоқ орман үлгісі, әсіресе бақылаулардың көбірек саны үшін жақсы дәлдік және Карра мәндерін көрсететін аңғал және үйренетін көпмүшелік логит үлгілеріне қарағанда айтарлықтай жақсарғанын көрсетеді.

Барлық пайдаланылған 16 алгоритмнің өнімділігін талдап және салыстыра отыра, RepTree алгоритмі TP жылдамдығы 0,902 және TF жылдамдығы 0,148 болатын ең жақсы болжамды берді. Ең нашар болжамы бар алгоритм TP көрсеткіші бар болғаны 0,790 болатын NaiveMulti болды. Дегенмен, егер біз ең нашар және ең жақсы нәтижелерді салыстыратын болсақ, TP жылдамдығының айырмашылығы шамамен 14% құрайтынын көреміз, бұл барлық алгоритмдердің жақсы орындалғанының және болжамның сенімділігінің көрсеткіші болып табылады. Сонымен қатар, барлық алгоритмдердің есептелген орташа TP жылдамдығы орташа баллдан жоғары орындалған 16 алгоритмнің 0,872 немесе 87,2% немесе 10 құрайды.



2-сурет. Шешімдер ағашының алгоритмдерінің көрсеткіші



3-сурет. Логистика алгоритмінің көрсеткіші

Логистика және Шешім ағашы алгоритмдері KNN және Bayes әдістерімен, 2-ші және 3-ші суреттерде көрсетілгендей, 0,871 пен тиісінше 0,832 салыстырғанда ең жоғары орташа TP жылдамдығын (0,889) санап берді.

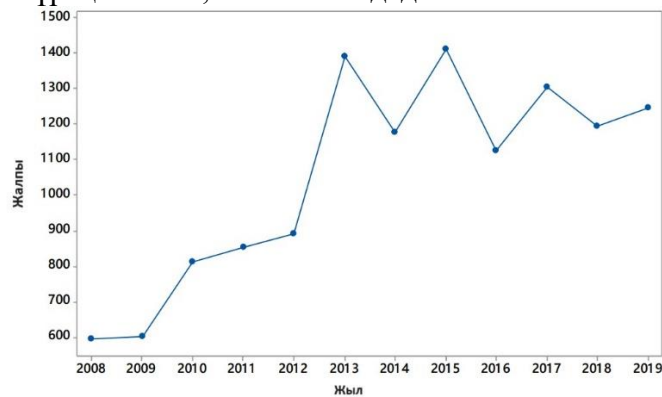
Өңделетін деректер алынғаннан кейін ол кіріс жиынына (70%) және валидация жиынына (30%) бөлінеді. Жаттығу жиынымен болжамды модель Шешімдер ағашы алгоритмі арқылы оқытылады және валидация жиынымен үлгі өнімділігі өнімділік, дәлдік, сезімталдық және т.б. көрсеткіштерге сәйкес бағаланады. Модель тексерілгеннен кейін болжам жасалады. Дезертирлеу қаупі бар-жоғын анықтау үшін жаңа студенттердің деректері өңделді. Алынған нәтижелер 2 кестеде жазылғандай, 90,43% сәйкес болжау көрсеткішін көрсетеді. Бұл ұсынылған модель сапасы мен тиімділігі жағынан сәйкес екенін білдіреді.

2-кесте. Болжам нәтижелері кестесі

Дәлдік	Айқындық	TP	TN	FN
--------	----------	----	----	----

90,43	95,91	93.48	59.26	6.52
-------	-------	-------	-------	------

4-сурет тұтастай алғанда студенттің нақты қызығушылығын тудыратын деректер сызбаларының уақытша қатары. График құбылмалы, бірақ тұрақты емес мәндерді көрсетеді. 2007/2008 жылдар кезеңінде жаңа студенттерді қабылдауға деген қызығушылық айтарлықтай өсті, ал 2012 жылы сюжет тұрақты емес, тіпті төмендеді.



4-сурет. XXX Университетінің жаңа студенттерінің уақыттық қатарының сюжеті

### Қорытынды

Әдебиетті шолуымызда студенттерді қабылдауды болжау барысында машиналық оқытуды пайдаланудың әртүрлі тәсілдері таңдалынды. Осы саладағы өзге де авторлардың түрлі жұмыстарын талдауда және оқуға түсуші студенттердің санын болжау үшін оңтайлы алгоритм айқындалды. Сондай-ақ, классикалық статистикалық әдістерді қолданудан бастап қазіргі заманғы машиналық оқыту алгоритмдеріне дейін осы мәселенің әртүрлі аспектілерін қамтитын әртүрлі авторлардың жұмыстары қарастырылды. Бұл алгоритмдер олардың әртүрлі нұсқаларының тиімділігі, сондай-ақ әртүрлі сценарийлер мен деректерге қолданылуы тұрғысынан қаралды.

Талдауымыздың нәтижесінде біз студенттерді қабылдауды болжаудың нақты әрі оңтайлы Шешім ағаштары алгоритмі деректердің нақты контекстіне және сипаттамаларына байланысты болуы мүмкін деген қорытындыға келдік. Тапсырманың ерекшелігі, білім беру жүйесінің ерекшеліктері және болжамның дәлдігіне қойылатын талаптарды ескеру маңызды орын алды.

### Қолданылған әдебиеттер тізімі

1. Qureshi, M. I., Khan, N., Raza, H., Imran, A., Ismail, F. Digital Technologies in Education 4.0. Does it Enhance the Effectiveness of Learning? A Systematic Literature Review // International Journal of Interactive Mobile Technologies. – 2021. – Vol. 15. – No. 04. – P. 31-47. – DOI:10.3991/ijim. v15i04.20291
2. Подорванюк Н. Ю. и др. Рейтинги университетов: кто виноват, кому выгодно и что делать? //Социодиггер. – 2021. – Т. 2. – №. 7 (12). – С. 62-69.
3. Ujkani B., Minkovska D., Stoyanova L. A machine learning approach for predicting student enrollment in the university //2021 XXX International Scientific Conference Electronics (ET). – IEEE, 2021. – С. 1-4.
4. Martínez-Garmendia J. Machine learning for product choice prediction //Journal of Marketing Analytics. – 2023. – С. 1-12.
5. Priyana E. D. Predictions of new students interest in entering the XXX University with statistic modeling //IOP Conference Series: Materials Science and Engineering. – IOP Publishing, 2020. – Т. 821. – №. 1. – С. 012030.
6. Vega H. et al. Intelligent system to predict university students dropout //vol. – 2022. – Т. 18. – С. 27-43.