

**Г.Т. Балақаева<sup>1</sup>, Д.К. Даркенбаев<sup>1\*</sup>, Д.С. Кулачар<sup>1</sup>**

<sup>1</sup>Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

\* e-mail: dauren.kadyrovich@gmail.com

## **DATA MINING ТЕХНОЛОГИЯСЫН ҚОЛДАНЫП ДЕРЕКТЕРДІ ӨНДЕУГЕ АРНАЛҒАН АҚПАРАТТЫҚ ЖҮЙЕ ӘЗІРЛЕУ**

### *Аңдатпа*

Мақалада DataMining технологиясын қолдана отырып, деректерді өңдеудің ақпараттық жүйесін жобалау және әзірлеу ұсынылған. Деректерді өңдеуде DataMining әдістері: логистикалық регрессия, кездейсоқ орман, шешім ағаштары, тірек векторлары әдісі, жасанды нейронды желі қолданылды және нәтижелері салыстырыла талданып кестеге салынып көрсетілді. Әзірленген ақпараттық жүйе құрылымданған, құрылымданбаған және жартылай құрылымданған емделуші азаматтардың деректерін өңдеу арқылы медицина қызметкерлеріне емделуші азаматтардың денсаулығына қатысты нақты болжам жасауға мүмкіндік береді. Машиналық оқыту алгоритмдерін пайдаланылып жобаланған ақпараттық жүйе, бірнеше жекелеген деректерді негізге алып өңдеу жұмыстарын жүргізеді. Дерекқордан алдыңғы деректерді алып, оларды талдайды, әрі қарай өңдейді ал жақсы нәтиже көрсеткен алгоритм нәтижесін салыстырып, жақсы нәтиже көрсеткен алгоритм таңдалады. Деректерді өңдеуге арналған ақпараттық жүйе қолданушыларға нақты уақыт режимінде болжам нәтижелерін алуға мүмкіндік береді. Авторлар құрылған ақпараттық жүйені өндіріске енгізуді және әрі қарай зерттеу жұмыстарын жалғастыруды жоспарлап отыр. Деректерді өңдеуге арналған ақпараттық жүйені білім саласында білім алушылардың білім деңгейлеріне болжамдар жасауда қолдануға болады.

*Түйін сөздер:* ақпараттық жүйе, деректер, алгоритм, өңдеу әдістер, data mining, машиналық оқыту.

**Г.Т.Балақаева<sup>1</sup>, Д.К.Даркенбаев<sup>1</sup>, Д.С.Кулачар<sup>1</sup>**

Казахский национальный университет им. Аль-Фараби, г. Алматы, Казахстан

## **РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ОБРАБОТКИ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ DATA MINING**

### *Аннотация*

В статье представлено проектирование и разработка информационной системы обработки данных с использованием технологии DataMining. При обработке данных использовались методы DataMining: логистическая регрессия, случайный лес, деревья решений, метод опорных векторов, искусственная нейронная сеть, результаты сравнивались, анализировались и представлялись в таблице. Разработанная информационная система за счет обработки структурированных, неструктурированных и полуструктурированных данных пациентов позволяет медицинским работникам делать точные прогнозы относительно состояния здоровья пациентов. Информационная система, разработанная с использованием алгоритмов машинного обучения, выполняет обработку на основе нескольких отдельных данных. Предыдущие данные берутся из базы данных, анализируются, далее обрабатываются, сравнивается алгоритм с лучшими результатами и выбирается алгоритм с лучшими результатами. Информационная система обработки данных позволяет пользователям получать результаты прогнозов в режиме реального времени. Авторы планируют запустить созданную информационную систему в производство и продолжать дополнять исследовательскую работу в дальнейшем. Информационная система обработки данных может быть использована для прогнозирования уровня знаний обучающихся в сфере образования.

*Ключевые слова:* информационная система, данные, алгоритмы, обработка, методы, datamining, машинное обучение.

G.T.Balakayeva<sup>1</sup>, D.K.Darkenbayev<sup>1</sup>, D.S.Kulachar<sup>1</sup>

<sup>1</sup>Al-Farabi Kazakh National University, Almaty, Kazakhstan

## DEVELOPMENT OF A DATA PROCESSING INFORMATION SYSTEM USING DATA MINING TECHNOLOGY

### Abstract

The article presents the design and development of a data processing information system using DataMining technology. When processing the data, DataMining methods were used: logistic regression, random forest, decision trees, support vector machine, artificial neural network, the results were compared, analyzed and presented in a table. The developed information system, by processing structured, unstructured and semi-structured patient data, allows medical professionals to make accurate predictions regarding the health status of patients. An information system developed using machine learning algorithms performs processing based on several individual data. Previous data is taken from the database, analyzed, further processed, the algorithm with the best results is compared, and the algorithm with the best results is selected. The data processing information system allows users to obtain forecast results in real time. The authors plan to put the created information system into production and continue to supplement the research work in the future. An information processing system can be used to predict the level of knowledge of students in the field of education.

*Keywords:* information system, data, algorithms, processing, methods, data mining, machine learning.

### Кіріспе

Қазіргі кезде деректердің өсуімен қатар оларды өңдеу де өзекті мәселелерге айналды. Технологияның дамуымен қатар интеллекталды жүйелерде дамып қоғамның қажеттілігін өтеп отыр. Түрлі салаларда болжам жасауға арналған ақпараттық жүйелер көбейіп, қолданушылардың қажеттілігіне қарай дамып жатыр. Мақаланың негізгі мақсатына сай Data Mining технологияларын қолданып, жеке адамдардың медициналық деректері негізінде қоғамда белең алған дерт қант диабетіне болжам жасалды. Құрылған ақпараттық жүйе басқада қолданыстағы жүйелер сияқты болжам жасауға арналған [1].

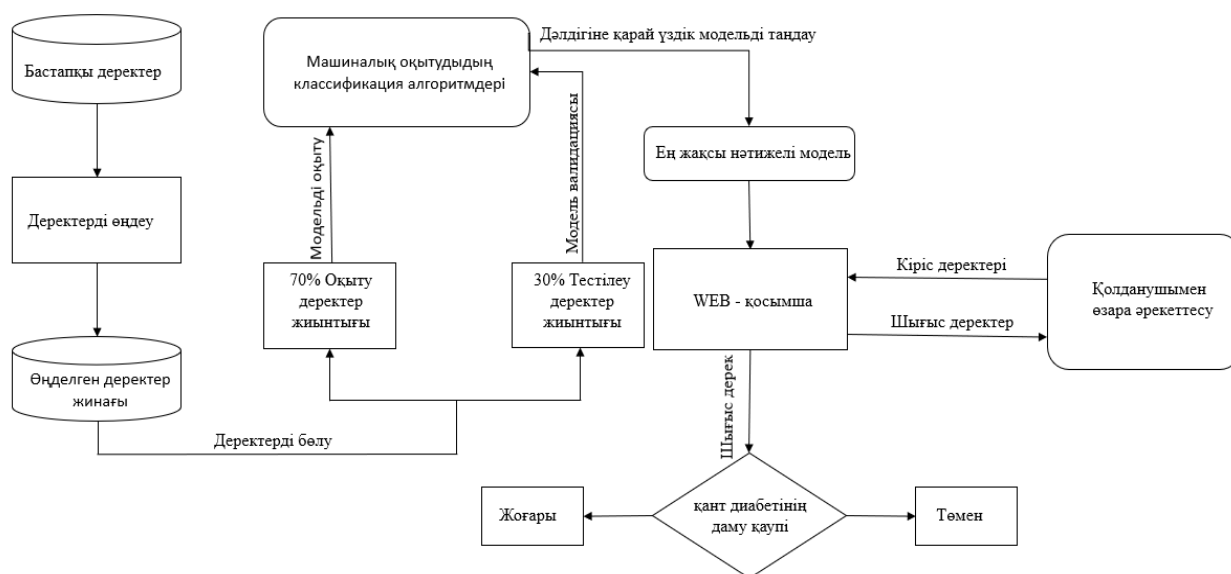
Қант диабеті-бұл бүкіл әлемдегі көптеген адамдарды қамтитын созылмалы метаболикалық ауру. Қант диабетін ерте анықтау және тиімді алдын алу оның таралуы мен ауырлығын төмендетуде маңызды рөл атқарады. Осыны ескере отырып, data mining әдістеріне негізделген болжамды модельдеу жүйелерін әзірлеу осы ауруды диагностикалау және емдеу саласында айтарлықтай пайда әкелуі мүмкін [2].

Бұл зерттеудің мақсаты деректерді data mining әдістерін қолдануға назар аудара отырып, нақты уақыттағы қант диабетін болжау медициналық жүйесінің архитектурасы мен әзірлеуін сипаттау болып табылады. Осы зерттеу аясында ұсынылған жүйені дамыту үшін кең контекст ұсыну үшін қант диабетін болжау жүйелері мен data mining әдістеріне арналған өзекті әдебиеттерге шолу жасалды. Деректер көздерін таңдауды, ақпаратты алдын ала өңдеу әдістерін және болжамды модельдеу тәсілдерін қоса алғанда, жүйені әзірлеу әдістемесі қарастырылады. Алынған болжамды модельдердің нәтижелері келтірілген және алынған нәтижелердің клиникалық практика үшін маңыздылығын көрсете отырып, қолданыстағы жүйелермен салыстырмалы талдау жүргізіледі. Нәтижесінде, жүргізілген зерттеу негізінде клиникалық тәжірибені одан әрі зерттеу және дамыту үшін маңызды қорытындылар мен ұсыныстар жасалады [3].

### Зерттеу әдіснамасы

*Ақпараттық жүйені жобалау және машиналық оқыту алгоритмдері.* Нақты уақыттағы болжам жасап, нақты нәтижелер алуға ұсынылған ақпараттық жүйе архитектурасы жүйенің элеуетті компоненттері мен жұмыс істеу принциптерін сипаттайтын құрылымдық тұжырымдама болып табылады. Бұл архитектура жүйені клиникалық тәжірибеде ыңғайлы және тиімді пайдалануды қамтамасыз етуге арналған. Жүйе пациенттерде қант диабетінің даму ықтималдығын болжау үшін қажетті әртүрлі деректерді жинауды және талдауды қамтиды. Деректер көздері демография, ауру тарихы және зертханалық нәтижелер туралы ақпаратты

қамтитын пациенттің электронды жазбалары болып табылады. Жүйе осы деректерге нақты уақытта қол жеткізуді қамтамасыз етеді және оларды алдын ала өңдеуді, соның ішінде жетіспейтін мәндерді жоюды, ауытқуларды өңдеуді және деректерді қалыпқа келтіруді жүзеге асырады[4]. Жүйеде болжам жасау үшін машиналық оқыту алгоритмдері қолданылатын болады. Бұл алгоритмдер әр пациенттің клиникалық деректеріне сүйене отырып, қант диабетінің даму ықтималдығын болжай алатын модель жасау үшін қолданылады. Ақпараттық жүйені жобалау үшін қолданылатын әдістемелер блок-схема түрінде төменде 1-ші суретте берілген.



Сурет 1. Ақпараттық жүйені жасау үшін қолданылатын әдістеменің блок-схемасы

Модельді алдыңғы деректер негізінде оқытуға болады, бұл жүйеге болжамдардың дәлдігі мен сенімділігін жақсартуға мүмкіндік береді. Жүйенің барлық компоненттері қауіпсіздік, құпиялылық және медициналық этика стандарттарына сәйкестік қағидаттарын ескере отырып әзірленуі керек. Жүйе медицина мамандарына онымен өзара әрекеттесуге және пациенттердегі қант диабетін диагностикалауға және басқаруға көмектесетін нақты уақыттағы болжамдар мен түсініктерді алуға мүмкіндік беретін интерфейсті қамтамасыз етеді. 1-ші суретте медициналық жүйенің жұмыс жасау блок-схемасы көрсетілген. Берілген құрылымдық тұжырымдама тек жалпы модельді білдіреді және одан әрі дамыту мен іске асыруды қажет етеді. Алайда, бұл нақты уақыт режимінде қант диабетін болжау жүйесінің дамуының ықтимал бағытын білдіреді.

Ақпараттық жүйелерді қолданып қант диабетін диагностикалау және болжау үшін қолдануға болатын машиналық оқыту алгоритмдерінің бірнеше түрі бар, соның ішінде [5]:

1. Логистикалық регрессия.
2. Кездейсоқ орман.
3. Тірек векторлары әдісі (SVM).
4. Шешімдер ағашы.
5. Жасанды нейронды желілер.

Data mining әдістері медициналық зерттеулер мен тәжірибеде ауруларды диагностикалау, болжау және емдеуді жоспарлау үшін кеңінен қолданылады. Қант диабеті жағдайында бірнеше зерттеулер пациенттер туралы ақпараттың үлкен массивтерін талдау және аурудың дамуы мен өршуінің тиісті қауіп факторларын анықтау үшін деректерді өндіру әдістерін қолдануды зерттеді. Khan және басқа зерттеушілер (2019) зерттеулерінің бірінде бірнеше демографиялық факторлар мен өмір салтына негізделген 2 типті қант диабетінің даму қаупін

болжау үшін шешім ағашы алгоритмі қолданылды. Нәтижелер шешім ағашының үлгісі қант диабетінің даму қаупін 80% дәлдікпен дәл болжай алатынын көрсетті. Басқа зерттеуде Zolbanin және басқа зерттеушілер. (2020) тірек векторлық машина алгоритмін қолдана отырып, қант диабетін диагностикалаудың болжамды моделі жасалды. Модель клиникалық және зертханалық айнымалылардың тіркесімін қолданды және нәтижелер тірек векторлық машинаның алгоритмі қант диабетін 84% сезімталдықпен және 87% спецификамен дәл анықтай алатынын көрсетті.

*Деректер жиынтығы.* Бұл мәліметтер жиынтығы Ұлттық қант диабеті және ас қорыту және бүйрек аурулары институтынан алынды. Мақсат - диагностикалық өлшеулер негізінде пациенттің қант диабеті бар-жоғын болжау. Бұл екілік (2-класс) классификация тапсырмасы. Әр класс үшін бақылаулар саны теңдестірілмеген. 8 кіріс айнымалысы және 1 шығыс айнымалысы бар 768 бақылау бар (сурет 2). Айнымалы атаулар келесідей.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Сурет 2. Қант диабеті туралы мәліметтер жиынтығы[6]

Pregnancies: Жүктілік саны

Glucose: глюкозаға төзімділік сынағы кезінде 2 сағаттан кейін қан плазмасындағы глюкозаның концентрациясы.

BloodPressure (қан қысымы): диастолалық қан қысымы (мм сын.бағ.) ст.)

SkinThickness: трицепс тері қатпарының қалыңдығы (мм)

Insulin: 2 сағаттық сарысулық инсулин (mu Ed/мл)

BMI: дене салмағының индексі (салмағы кг / (биіктігі м)<sup>2</sup>)

DiabetesPedigreeFunction: отбасылық тарих негізінде 2 типті қант диабетінің даму қаупін анықтайтын функция; функция неғұрлым көп болса, 2 типті қант диабетінің даму қаупі соғұрлым жоғары болады.

Age: жасы (жылдар)

Outcome: класс айнымалысы (0 немесе 1)

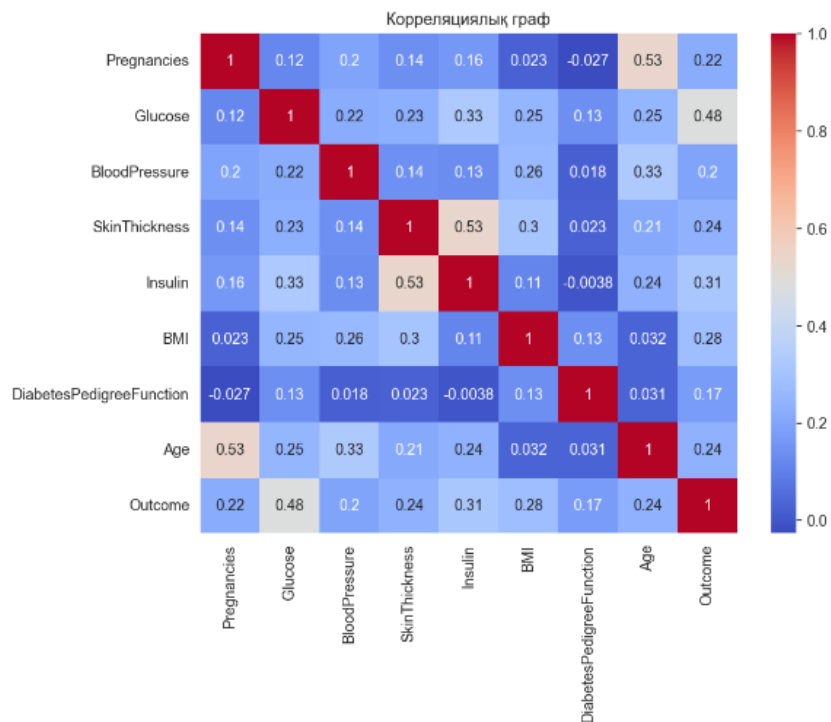
Ақпараттық жүйені қолданып қант диабетінің даму ықтималдығын болжау үшін, шешім ағашы алгоритмі қолданылады. Бұл алгоритм қант диабетінің ең маңызды болжаушыларын анықтауға және оларды клиникалық дәлелдеріне сүйене отырып, белгілі бір пациенттің қаупін болжау үшін қолдануға мүмкіндік береді.

*Деректерді талдау.* Деректерді талдаудың жаңа жобасын бастағанда, деректерді тексеру, түсіну және тазарту маңызды [7]. Деректердің алғашқы он жолын зерттегенде, инсулин деңгейі үшін де, терінің қалыңдығы үшін де өлшемдер нөлге тең болады. Егер пациенттің инсулин деңгейі мен терінің қалыңдығы нөлге тең болса, бұл үлкен медициналық мәселе болар еді. Осылайша, осы деректер жиынтығында нөл саны жетіспейтін немесе нөлдік деректерді көрсету үшін қолданылады деген қорытынды жасауға болады. Мұнда біз жолдардың жартысында деректер жетіспейтін бағандар бар екенін көреміз. Дегенмен, ешкім айтпайтын немесе жиі айтпайтын бір нәрсе бар - терінің қалыңдығы мен инсулин деңгейі үшін өткізіп алған мәндердің жоғары жиілігі, терінің қалыңдығы үшін 215 мән (28,8%), ал инсулин үшін 359 өткізіп алған мән (48,2%). Түсінікті болу үшін жетіспейтін мәндер нөлдер ретінде жазылады.

Деректерді алдын ала өңдеу. Жетіспейтін мәндерді келесі жолдармен өңдеуге болады [8]:

- жетіспейтін мәндерді осы бағанның орташа немесе медианасымен ауыстыру.
- өткізіп алған мәндердің атрибуттарын жою.
- жоқ мәндерді нөлге ауыстырыңыз.

Мұнда бірінші әдіс қарастырылады, өйткені сапалы атрибуттарды жою маңызды деректердің ықтимал жоғалуына әкеледі, ал жетіспейтін қасиеттерді нөлдермен ауыстыру нашар жіктеуге әкеледі. Жетіспейтін мәндер жаһандық деректер жиынындағы сәйкес атрибуттардың медианалық мәнімен ауыстырылады. Деректерді өңдеп болғаннан кейінгі корреляциялық матрица (сурет 3) бағандар арасындағы қатынастарды көре аламыз.



Сурет 3. Деректер жиынтығының корреляциялық матрицасы[9]

### Зерттеу нәтижелері

Деректер жиынтығы демографиялық ақпаратты, клиникалық өлшемдерді және зертханалық зерттеулердің нәтижелерін қамтыды. Біз қант диабеті қаупін болжау үшін бес машиналық оқыту алгоритмін қолдандық: логистикалық регрессия, шешім ағашы, кездейсоқ орман, тірек векторлық машина және нейрондық желі. Біз әр алгоритмді 70% деректер жиынтығында оқыттық және оның тиімділігін 30% деректер жиынтығында бағаладық. Біз әрбір алгоритмді деректер жиынында үйреттік және олардың тиімділігін accuracy, precision және F1 score сияқты көрсеткіштер арқылы бағаладық. Біз сондай-ақ қант диабеті қаупін болжау үшін ең маңызды айнымалыларды анықтау үшін белгілерді таңдадық.

Шешімдер ағашы. шешімдердің ағаш үлгісін және олардың салдарын, соның ішінде оқиғалардың кездейсоқ нәтижелерін, ресурстардың шығындарын және пайдалылығын пайдаланатын шешімдерді қолдау құралы [10]. Бұл тек шартты басқару операторларын қамтитын алгоритмді көрсетудің бір жолы. Сонымен қатар, шешім ағашы-бұл блок-схемаға ұқсас құрылым, онда әрбір ішкі түйін атрибуттың "сынағы" болып табылады және әрбір тармақ Сынақ нәтижесін білдіреді. Әрбір жапырақ түйіні сынып белгісін білдіреді (қант диабеті мысалындағыдай барлық атрибуттарды есептегеннен кейін қабылданған шешім). Тамырдан жапыраққа дейінгі жолдар жіктеу ережелерін білдіреді. Шешім ағашын жіктеу үшін де, регрессия үшін де қолдануға болады. Біздің тәуелді "нәтиже" айнымалысы жіктеу

айнымалысы болғандықтан, категориялық шешім ағашы әдісін қолдану Пима үндістерінің деректер жиынтығы үшін қолайлы алгоритмдердің бірі болды.

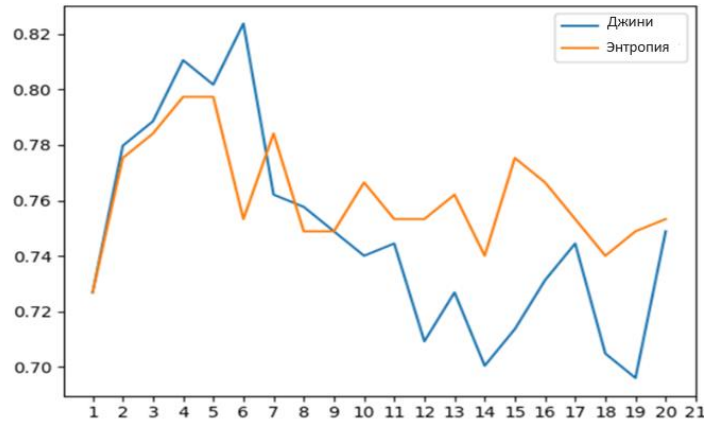
Шешім ағашы алгоритмін қолданғанды ағаш қай метриkanı қолданғанда дәлдік жоғары болатынын анықтау үшін ағаштың тереңдігі мен gini және entropy метрикаларының арасындағы тәуелділікті анықтап көрдік. “Giniimpurity” метрикасы (1) C класстарының жиынтығы және берілген Q деректер жиынтығы үшін:

$$G(Q) = \sum_{c \in C} p_c(1-p_c) \quad (1)$$

$p_c$ — C классының пайда болу ықтималдығы (2).  $N_Q$

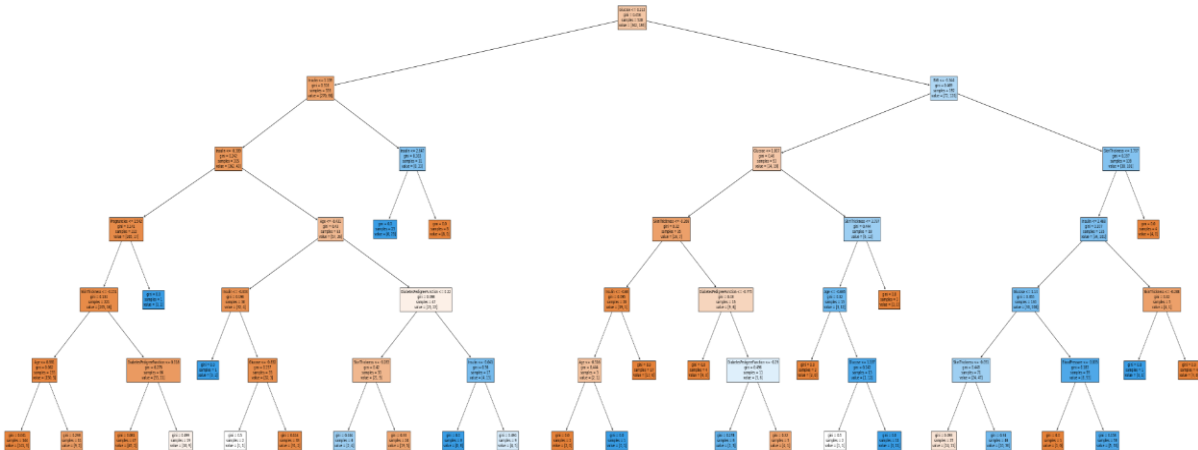
$$p_c \frac{1}{N_Q} \sum_{x \in Q} y_{class} = c \quad (2)$$

Нәтижелер көрсеткендей, gini метрикасы entropy метрикасынан асып, ағаштың тереңдігі 6 болғанда шамамен 82% дәлдікке қол жеткізді, ал entropy метрикасы ағаштың тереңдігі 5-ке тең болғанда шамамен 80% дәлдікке жетті және болжау дәлдіктері төменде 4-ші суретте берілген.



Сурет 4. Шешім ағашы алгоритмінің ағаш тереңдігіне байланысты болжау дәлдігі

Бұл нәтижелер шешім ағашы алгоритмдерінде метриkanı таңдаудың маңыздылығын көрсетеді және қант диабетін болжау саласындағы зерттеушілер мен тәжірибешілерге құнды ақпарат береді. Шешім ағашының визуалды түрі төменде 5-ші суретте берілген.



Сурет 5. Шешімдер ағашының визуалды түрдегі көрсетілімі [11]

Ақпараттық жүйені қолдану арқылы пациенттердің деректерін енгізу және машиналық оқыту алгоритмдерін қолдану арқылы қант диабетінің даму қаупіне болжам нәтижелері төменде 1-ші кестеде берілген.

Кесте 1. Болжау алгоритмдерінің көрсеткіштер бойынша салыстырмалы тиімділігі

Модель/Метрика	Шығыс мәні	Дәлдік	Кері шақыру	f1-ұпайы	Қолдау көрсету	Дәлдік
Шешім ағашы	0	0.85	0.88	0.87	152	81.9%
	1	0.74	0.69	0.72	75	
Логистикалық регрессия	0	0.78	0.88	0.82	152	74.8%
	1	0.66	0.49	0.56	75	
Кездейсоқ орман	0	0.82	0.87	0.85	152	77.9%
	1	0.70	0.63	0.66	75	
Тірек векторлары әдісі	0	0.80	0.75	0.86	152	78%
	1	0.69	0.68	0.63	75	
Жасанды нейронды желілер	0	0.81	0.80	0.80	152	75.2%
	1	0.65	0.68	0.66	75	

Жүйе нақты уақыт режимінде жұмыс істейтін болады, бұл медицина қызметкерлеріне қауіптілігі жоғары пациенттерді жедел анықтауға және тиісті алдын алу шараларын қабылдауға мүмкіндік береді. Осы зерттеуде сипатталған медициналық жүйе әлі әзірленбеген және сынақтан өтпегенімен, оны қолдану қант диабетінің диагностикасы мен болжамын айтарлықтай жақсартып, осы созылмалы ауруды тиімдірек басқаруға әкелуі мүмкін деп болжануда.

### Дискуссия

Соңғы жылдары зерттеушілердің назары машиналық оқыту алгоритмдері негізінде болжамдар жасауға аударылды [12]. Мұндай жүйелерде деректерді талдаудың әртүрлі әдістері қолданылады, соның ішінде шешім ағаштары, жасанды нейрондық желілер, логистикалық регрессия және тірек векторлық машиналар. Бұл әдістердің мақсаты қант диабетінің ең маңызды болжаушыларын анықтау және оларды белгілі бір науқаста қант диабетінің даму ықтималдығын болжау үшін пайдалану болып табылады.

Бірқатар зерттеулер сонымен қатар денсаулық сақтау мамандарына пациенттің нақты уақыттағы жағдайы туралы деректерді беру үшін глюкозаны үздіксіз бақылау сияқты нақты уақыттағы бақылау жүйелерін пайдалануды зерттеді. Мысалы, зерттеуде Charleer және басқа зерттеушілер (2018) 1 типті қант диабетімен ауыратын науқастар үшін глюкозаның үздіксіз мониторингін инсулин сорғысы терапиясымен біріктіретін нақты уақыттағы бақылаудың ақпараттық жүйесі жасалды. Нәтижелер жүйенің глюкозаны бақылауды жақсартқанын және гипогликемиялық оқиғалардың жиілігін төмендететінін көрсетті [13].

Deu және басқа зерттеушілер қант диабеті ауруларын болжауға арналған веб-қосымшаны ұсынды. Тірек векторлық машина (SVM), K – nearest neighbors (KNN), Naive Bayes, логистикалық регрессия және жасанды нейрондық желі (ANN) сияқты әртүрлі Машиналық оқыту алгоритмдері қолданылды. Бұл зерттеу жасанды нейрондық желімен (ANN) біріктірілген Min-Max масштабтау дәлдігі әдісі ең жоғары дәлдікке ие екенін көрсетті [14].

Kumarі және басқа зерттеушілер, ауруларды жіктеу және диагностикалау моделін ұсынды, әр түрлі кішірейту алгоритмдері мен жіктеу алгоритмдері қолданылды. Оның денсаулық сақтауда қолдану аясы кең, мысалы, медициналық бейнелер алу, медициналық білім алу, медициналық шешімдерді қолдау, пациенттерді жалпы басқару және ақуыз-ақуыздың өзара әрекеттесуі.

Chauhan және басқа зерттеушілер веб-формада қант диабетін болжау үшін үш жіктеу моделін қолданды. Бұл модельдер шешім ағашынан, нейрондық желіден және байестен тұрды. Әр модельдің сенімділігін тексеру үшін дәлдік пен ROC қисығы есептелді, олар басқалармен салыстырылды [14].

### Қорытынды

Data mining әдістерін қолдана отырып, нақты уақыт режимінде ауру түрлерін болжаудың ақпараттық жүйесін жобаланды. Машиналық оқыту алгоритмдерін қолданып түрлі ауруларға болжам жасау тиімділігін едәуір арттыру қазіргі заманғы өзекті мәселелердің бірі болып отыр. Зерттеу нәтижелері машиналық оқыту алгоритмдерінің жоғары дәлдігін көрсетеді, бұл құрылған ақпараттық жүйені қолданып науқастарға медициналық көмек көрсету сапасын жақсарту мүмкіндігінің жоғары екенін көрсетеді. Дегенмен, бұл салада нақты болжамдар әлі де үлкен зерттеулерді қажет етеді. Мақала авторлары машиналық оқыту алгоритмдері негізінде түрлі болжамдар жасау жұмыстары бойынша тәжірибелері бар және зерттеу жұмыстарын әрі қарай жалғастыруды жоспарлауда.

Тұтастай алғанда, Data Mining технологиясы негізінде әзірленген ақпараттық жүйені нақты уақыт режимінде медицина саласында қант диабеті ауруына болжам жасап, алдын алу мақсатында қолдануға болады. Ақпараттық жүйені қолдану арқылы жасалған болжамдар науқастардың ауруларының алдын алуға және емдеу процестерінің тиімділігін айтарлықтай арттыруы мүмкін. Авторлардың пайымдауынша аталмыш ақпараттық жүйені медицина, білім, үлкен көлемді құрылымданбаған деректерді өңдеуде және несиелік скорингте болжамдар жасауға пайдалануға болады.

### Пайдаланылған дереккөздер тізімі

- [1] *Machine Learning-Based Application for Predicting Risko fType 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study* / A.Syed; T. Khan - *IEEE Access* 30 October 2020 <https://ieeexplore.ieee.org/document/9245498>
- [2] *A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and Deploying a model ensemble* / M. Zolbanin, S. Piri, D. Delen, T. Liu - *Decision Support Systems Volume 101*, September 2017, Pages 12-27 <https://www.sciencedirect.com/science/article/abs/pii/S0167923617300908>
- [3] *Quality of Life and Glucose Control After 1 Year of Nationwide Reimbursement of Intermittently Scanned Continuous Glucose Monitoring in Adults Living With Type 1 Diabetes (FUTURE): A Prospective Observational Real-World Cohort Study* / S. Charleer, B. Broos, S. Fieuws - *Diabetes Care* 2020;43(2):389–397 <https://diabetesjournals.org/care/article/43/2/389/36133/Quality-of-Life-and-Glucose-Control-After-1-Year>
- [4] *Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm* / S. Dey, A. Hossain, Md. Rahman - *Published in 2018 21st International Conference of Computer and Information Technology (ICIT)* <https://ieeexplore.ieee.org/document/8631968>
- [5] *Classification Of Diabetes Disease Using Support Vector Machine* / V. Anuja Kumari, R.Chitra - *March - April 2013 International Journal of Engineering Research and Applications*
- [6] *Improve Classification Performance In Diabetes Prediction* / C. Chauhan, S. Karvande - [http://www.oaijse.com/VolumeArticles/FullTextPDF/449\\_8.Improve\\_classification\\_performance\\_in.pdf](http://www.oaijse.com/VolumeArticles/FullTextPDF/449_8.Improve_classification_performance_in.pdf)
- [7] *Real-time crash prediction on freeways using data mining and emerging techniques* / J. You, J. Wang, J. Guo - *Journal of Modern Transportation volume 25*, pages 116–123 (2017) <https://link.springer.com/article/10.1007/s40534-017-0129-7>
- [8] Кондрашов Ю.Н. *Анализ данных и машинное обучение на платформе MS SQL Server: учебное пособие* / Ю.Н. Кондрашов. – Москва : РУСАЙНС, 2020. – 304 б.
- [9] Robert L. *Learning Data mining with Python.* / L. Robert. – Birmingham : Packt Publishing, 2015. – 317 б.
- [10] Burns A. *Real-Time Systems and Programming Language Ada, Real-Time Java and C, Real-Time POSIX* / A. Burns, A. - Wellings University of Yor, 2009. – 602 б.
- [11] Williams R. *Real-Time Systems Development* / R. Williams - Waltham: Elsevier, 2006. – 450 б.
- [12] Ian H. *Data mining Practical Machine Learning Tools and Techniques.* / H. Ian, F. Eibe, H. Mark, P. Christopher – Изд. 4-е – Cambridge : Todd Green, 2017. – 622 б.
- [13] Jared D. *Big Data, Data mining and Machine Learning* / D. Jared Hoboken : John Wiley & Sons, 2014. – 265 б.
- [14] Darius M. *Data mining for genomics and proteomics Analysis of Gene and Protein Expression Data* / M. Darius - Hoboken: John Wiley & Sons, 2010. – 349



References

- [1] *Machine Learning-Based Application for Predicting Risk of Type 2 Diabetes Mellitus (T2DM) in Saudi Arabia: A Retrospective Cross-Sectional Study* / A. Syed; T. Khan - *IEEE Access* 30 October 2020 <https://ieeexplore.ieee.org/document/9245498>
- [2] *A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble* / M. Zolbanin, S. Piri, D. Delen, T. Liu - *Decision Support Systems Volume 101*, September 2017, Pages 12-27 <https://www.sciencedirect.com/science/article/abs/pii/S0167923617300908>
- [3] *Quality of Life and Glucose Control After 1 Year of Nationwide Reimbursement of Intermittently Scanned Continuous Glucose Monitoring in Adults Living With Type 1 Diabetes (FUTURE): A Prospective Observational Real-World Cohort Study* / S. Charleer, B. Broos, S. Fieuws - *Diabetes Care* 2020;43(2):389–397 <https://diabetesjournals.org/care/article/43/2/389/36133/Quality-of-Life-and-Glucose-Control-After-1-Year>
- [4] *Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm* / S. Dey, A. Hossain, Md. Rahman - *Published in 2018 21st International Conference of Computer and Information Technology (ICCIT)* <https://ieeexplore.ieee.org/document/8631968>
- [5] *Classification Of Diabetes Disease Using Support Vector Machine* / V. Anuja Kumari, R. Chitra - *March - April 2013 International Journal of Engineering Research and Applications*
- [6] *Improve Classification Performance In Diabetes Prediction* / C. Chauhan, S. Karvande - [http://www.oaijse.com/VolumeArticles/FullTextPDF/449\\_8.Improve\\_classification\\_performance\\_in.pdf](http://www.oaijse.com/VolumeArticles/FullTextPDF/449_8.Improve_classification_performance_in.pdf)
- [7] *Real-time crash prediction on freeways using data mining and emerging techniques* / J. You, J. Wang, J. Guo - *Journal of Modern Transportation* volume 25, pages 116–123 (2017) <https://link.springer.com/article/10.1007/s40534-017-0129-7>
- [8] *Kondrashov Yu.N. (2020) Analiz dannykh i mashinnoe obuchenie na platforme MS SQL Server [Data analysis and machine learning on the MS SQL Server platform: textbook]. Yu.N. Kondrashov. Moscow: RUSAINS, 304. (In Russian)*
- [9] *Robert L. Learning Data mining with Python.* / L. Robert. – Birmingham : Packt Publishing, 2015. – 317 б.
- [10] *Burns A. Real-Time Systems and Programming Language Ada, Real-Time Java and C, Real-Time POSIX* / A. Burns, A. - *Wellings University of Yor*, 2009. – 602 б.
- [11] *Williams R. Real-Time Systems Development* / R. Williams - *Waltham: Elsevier*, 2006. – 450 б.
- [12] *Ian H. Data mining Practical Machine Learning Tools and Techniques.* / H. Ian, F. Eibe, H. Mark, P. Christopher – *Изд. 4-e* – *Cambridge : Todd Green*, 2017. – 622 б.
- [13] *Jared D. Big Data, Data mining and Machine Learning* / D. Jared – *Hoboken : John Wiley & Sons*, 2014. – 26 p.
- [14] *Darius M. Data mining for genomics and proteomics Analysis of Gene and Protein Expression Data* / M. Darius - *Hoboken: John Wiley & Sons*, 2010. – 349.