

HISOBLASH VA AMALIY MATEMATIKA MUAMMOLARI

ПРОБЛЕМЫ ВЫЧИСЛИТЕЛЬНОЙ
И ПРИКЛАДНОЙ МАТЕМАТИКИ

PROBLEMS OF COMPUTATIONAL
AND APPLIED MATHEMATICS



ПРОБЛЕМЫ ВЫЧИСЛИТЕЛЬНОЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ

№ 1(46) 2023

Журнал основан в 2015 году.

Издается 6 раз в год.

Учредитель:

Научно-исследовательский институт развития цифровых технологий и
искусственного интеллекта.

Главный редактор:

Равшанов Н.

Заместители главного редактора:

Азамов А.А., Арипов М.М., Шадиметов Х.М.

Ответственный секретарь:

Ахмедов Д.Д.

Редакционный совет:

Азамова Н.А., Алоев Р.Д., Бурнашев В.Ф., Загребина С.А. (Россия),
Задорин А.И. (Россия), Игнатъев Н.А., Ильин В.П. (Россия),
Исмагилов И.И. (Россия), Кабанихин С.И. (Россия), Карачик В.В. (Россия),
Курбонов Н.М., Маматов Н.С., Мирзаев Н.М., Мухамедиева Д.Т., Назирова Э.Ш.,
Нормуродов Ч.Б., Нуралиев Ф.М., Опанасенко В.Н. (Украина), Раджабов С.С.,
Расулов А.С., Садуллаева Ш.А., Самаль Д.И. (Беларусь),
Старовойтов В.В. (Беларусь), Хаётов А.Р., Хамдамов Р.Х., Хужаев И.К.,
Хужаеров Б.Х., Чье Ен Ун (Россия), Шабозов М.Ш. (Таджикистан),
Шадиметов Х.М., Dimov I. (Болгария), Li Y. (США), Mascagni M. (США),
Min A. (Германия), Rasulev V. (США), Schaumburg H. (Германия), Singh D. (Южная
Корея), Singh M. (Южная Корея).

Журнал зарегистрирован в Агентстве информации и массовых коммуникаций при
Администрации Президента Республики Узбекистан.

Регистрационное свидетельство №0856 от 5 августа 2015 года.

ISSN 2181-8460, eISSN 2181-046X

При перепечатке материалов ссылка на журнал обязательна.

За точность фактов и достоверность информации ответственность несут авторы.

Адрес редакции:

100125, г. Ташкент, м-в. Буз-2, 17А.

Тел.: +(99871) 231-92-45.

E-mail: journals@airi.uz.

Сайт: journals.airi.uz (www.pvpm.uz).

Дизайн и компьютерная вёрстка:

Шарипов Х.Д.

Отпечатано в типографии НИИ РЦТИИ.

Подписано в печать 28.02.2023 г.

Формат 60x84 1/8. Заказ №1. Тираж 100 экз.

Содержание

<i>Равшанов Н., Аминов С.М.</i> Моделирование многофазной фильтрации в многослойной деформируемой пористой среде	5
<i>Ибрагимов А.А., Хамроева Д.Н.</i> Полная проблема собственных значений для несимметричных интервальных матриц	31
<i>Жалолов О.И., Хаятов Х.У.</i> Алгоритм построения оптимальной интерполяционной формулы в пространстве Соболева $\tilde{W}_2^{(m)}(T_1)$	47
<i>Равшанов Н., Мухамедиева Д.Т., Курбонов Н.М., Тухтамуродов Н.У.</i> Моделирование нелинейной фильтрации флюидов в пористой среде с применением технологий искусственного интеллекта	60
<i>Мадражимов Ш,Ф., Махаров К.Т.</i> Классификация объектов выборки с пропусками в данных	78
<i>Мухамедиева Д.Т., Рустамов Е.Н.</i> Алгоритм обработки знаний	88
<i>Адылова Ф.Т., Давронов Р.Р., Сафаров Р.А., Кушимуратов С.И.</i> Теория чат-ботов и её приложения в здравоохранении	101
<i>Равшанов Н., Пекось О.А., Бакаев И.И.</i> Прогнозирование сердечно-сосудистых заболеваний методами машинного обучения	109
<i>Исмаилов О.М., Мирзахалилов С., Исмаилов М.О.</i> Исследование методов и алгоритмов репликации в системах с распределенной базой данных	116

УДК 004.4

ПРОГНОЗИРОВАНИЕ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

¹**Равшанов Н.*, ¹*Пекось О.А.*, ²*Бакаев И.И.*

*ravshanzade-09@mail.ru

¹Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта,
100125, Узбекистан, Ташкент, м-в. Буз-2, 17А.;
²Бухарский государственный университет,
200118, Узбекистан, Бухара, ул. М. Икбол дом 11.

Начиная со второй половины прошлого века, сердечно-сосудистые заболевания прочно занимают верхние строки в структуре смертности по всему миру, а в некоторых странах – более двух третей от общего количества смертей ежегодно. В этой связи, крайне актуальной проблемой является их ранняя диагностика, что многократно повышает успех лечения и вероятность положительного исхода заболеваний. Традиционные методики оценки риска сердечно-сосудистых заболеваний неявно допускают линейную взаимосвязь факторов риска с исходом сердечно-сосудистых заболеваний. Кроме того, обычно риск сердечно-сосудистых заболеваний рассчитывается только с использованием клинической информации без учета прочих важных факторов, например, социальных детерминант здоровья. Так как машинное обучение уже давно стало общепринятым средством обнаружения закономерностей в данных, то данный класс методов искусственного интеллекта, сегодня, активно применяется в клинической кардиологии. В том числе при решении задач: автоматического распознавания и расшифровки цифровых сигналов ЭКГ, ЭхоКГ, ХМ ЭКГ; диагностики патологий; автоматической классификации больных по различным признакам; оценки рисков развития неблагоприятных событий сердечно-сосудистой системы; автоматического подбора тактик лечения; прогнозирования результатов кардиологической реабилитации и т.д. Цель данной работы состояла в решении задачи по оценке склонности пациентов к сердечно-сосудистым заболеваниям путем применения одного из популярных методов машинного обучения — наивного байесовского классификатора к набору данных амбулаторного обследования пациентов. Алгоритм был реализован на языке python с использованием библиотеки scikit-learn.

Ключевые слова: сердечно-сосудистые заболевания, алгоритмы машинного обучения, наивный байесовский классификатор.

Цитирование: *Равшанов Н., Пекось О.А., Бакаев И.И.* Прогнозирование сердечно-сосудистых заболеваний методами машинного обучения // Проблемы вычислительной и прикладной математики. – 2023. – № 1(46). – С. 109-115.

1 Введение

На сегодняшний день сердечно-сосудистые заболевания (ССЗ), несмотря на значительный прогресс в их диагностике и лечении, являются одними из основных причин смерти населения по всему миру. Это возлагает большую ответственность на медперсонал. Традиционно, для определения заболевания проводится диагностика на основе информации, полученной от ряд разнообразных анализов крови или различных датчиков который прикоплен на теле пациента. Из-за сложности заболевания его трудно обнаружить по многим симптомам, таким как высокое кровяное давление,

уровень холестерина и аномальная частота пульса. Появлением методов искусственного интеллекта, таких как машинное обучение, во всех сферах медицины произошли большие изменения включая в кардиологии. В результате появилась эффективность прогнозирования заболеваний на основе больших массивов медицинских данных. Алгоритмы машинного обучения могут радикально изменить медицинскую практику, предлагая медперсоналу новые инструменты для интерпретации данных и принятия клинических решений. В этом плане проводится исследование для прогнозирования раннее выявление сердечно-сосудистых заболеваний, который имеет решающее значение для спасения жизни пациентов. Прогнозирование в значительной степени зависит от методов классификации. Эти методы помогают определить, нуждается ли пациент в лечении или склонность пациента к заболеванию. Существует множества работ по применению алгоритмов машинного обучения таких как наивный Байес, дерева решений, регрессия и т.п. Далее рассмотрим некоторых из них.

2 Связанные работы

Авторы [1] оценили общую способность алгоритмов машинного обучения для прогнозирования сердечно-сосудистых заболеваний. В исследовании была использовано различные базы данных, опубликованных в 2019 году, которые могут предсказывать такие заболевания, как ишемическая болезнь сердца, сердечные аритмии, сердечная недостаточность и инсульт. При анализе прогноза использовалась метрика площади под кривой.

Исследователями [2] было изучено связь между концентрации металлов в крови и моче, сердечно-сосудистыми заболеваниями и смертностью от рака. В качестве источника была использовано наборы данных из Национального обследования состояния здоровья и питания. Регрессия Пуассона использовалась для изучения воздействия как на один, так и на несколько металлов.

Авторами [3] проведён оценка алгоритмов машинного обучения для прогнозирования прогрессирования атеросклероз сосудов сердца. Оценка модели сравнивалась с оценкой риска сердечно-сосудистого атеросклероза. Авторы утверждают, что оценка незамеченных смещений в наборе данных с использованием машинного обучения по-прежнему остается сложной задачей.

Группа исследователей [4] оценили алгоритмы машинного обучения для прогнозирования разных заболеваний в кардиологии таких как ишемической болезни сердца, сердечной недостаточности, инсульта и сердечные аритмии. Для прогнозирования ишемической болезни сердца алгоритм бустинг имел объединенную площадь под кривой (AUC) 0,88 (95% Доверенный интервал 0,84–0,91), а специально разработанные алгоритмы имели объединенную AUC 0,93 (95% Доверенный интервал 0,85–0,97). Для прогнозирования инсульта алгоритмы машины опорных векторов (SVM) имели объединенную AUC 0,92 (95% ДИ 0,81–0,97), алгоритмы бустинг имели объединенную AUC 0,91 (95% ДИ 0,81–0,96), а сверточная нейронная сеть (CNN) алгоритмы имели объединенную AUC 0,90 (95% ДИ 0,83–0,95). Несмотря на неадекватность исследований каждого алгоритма метааналитической методологии как для сердечной недостаточности, так и для сердечных аритмий, поскольку доверительные интервалы между различными методами перекрываются, не показывая различий, SVM может превзойти другие алгоритмы в этих областях.

В работе [5] предложен метод прогнозирования состояния сердца, основанный на IoT и машинном обучении. Данные, собранные с человеческого тела, были нормализованы перед тем, как их использовали алгоритмы машинного обучения для расчета

и прогнозирования общего состояния сердца пациента, результаты оказались вполне удовлетворительными.

В исследовании [6] основное внимание уделяется сравнению традиционных подходов, таких как логистическая регрессия, К-ближайшие соседи (KNN), наивный байесовский метод (NB), метод опорных векторов (SVM), нейронные сети (NN) и предлагаемая модель прогнозирования CNN. Набор данных репозитория машинного обучения UCI для экспериментов и прогнозов сердечно-сосудистых заболеваний (ССЗ) с точностью 94 %.

Авторами [7] исследование предлагается ансамблевый подход, в котором используются модели машинного обучения (ML) и глубокого обучения (DL) для прогнозирования вероятности развития у человека сердечно-сосудистых заболеваний. В исследовании использовано шесть алгоритмов классификации для прогнозирования сердечно-сосудистых заболеваний. Модели обучаются с использованием общедоступного набора данных о случаях сердечно-сосудистых заболеваний. Для извлечения важных признаков сердечно-сосудистых заболеваний используется случайный лес (RF). В результаты эксперимента модель ансамбля ML достигает наилучшей точности прогнозирования заболеваний 88,70%.

Исследователями [8] предложен множество жизнеспособных подходов к выявлению сердечных заболеваний с помощью машинного обучения. Следовательно, были опубликованы исследования по разработке медицинских приложений с использованием различных алгоритмов и подходов машинного обучения.

В работе [9] создан метод повышения точности прогнозирования сердечно-сосудистых заболеваний. Предложенный смешанный метод Random Forest и линейной модели был точен на 88,7% в прогнозировании сердечно-сосудистых заболеваний.

Учёными [10] представлен гибридный подход к диагностике сердечных заболеваний. Результат показывает, что логистическая регрессия имеет точность 89% при прогнозировании сердечных заболеваний.

Авторы [11] предложили гибридную модель для прогнозирования ишемической болезни сердца. Для создания модели они использовали три алгоритма машинного обучения: RF, DT и их комбинацию. Гибридная модель оказалась самой точной с точностью 88,7%.

Вышеуказанные исследования доказывают, что применение алгоритмов машинного обучения даёт возможность определения склонности пациента к заболеваниям и принятие клинических решений таких как нуждается ли пациент в лечении. Для демонстрации используем классификатор Наивного Байеса.

3 Постановка задачи

Дано множества: $O = \{o_1, o_2, \dots, o_j, \dots, o_n\}$, где o_j - объект, который содержит информацию о пациентах с кардиологическими проблемами [12] (таблица 1). Описание полей таблицы 1 приведено в таблице 2.

Каждый o_j описывается набором переменных:

$$o_j = \{x_1, x_2, \dots, x_h, \dots, x_m, y\},$$

где x_h — независимые переменные (**age, gender, height, weight, ap_hi, ap_lo, cholesterol, gluc, smoke, alco, active**), значения, которые известны и на основании которых определяется значение зависимой переменной y (**cardio**). Набор независимых переменных обозначают в виде вектора:

$$X = \{x_1, x_2, \dots, x_h, \dots, x_m\},$$

Таблица 1

id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	17474	1	156	56.0	100	60	1	1	0	0	0	0
8	21914	1	151	67.0	120	80	2	2	0	0	0	0
9	22113	1	157	93.0	130	80	3	1	0	0	1	0
12	22584	2	178	95.0	130	90	3	3	0	0	1	1
13	17668	1	158	71.0	110	70	1	1	0	0	1	0
14	19834	1	164	68.0	110	60	1	1	0	0	0	0
15	22530	1	169	80.0	120	80	1	1	0	0	1	0
16	18815	2	173	60.0	120	80	1	1	0	0	1	0
18	14791	2	165	60.0	120	80	1	1	0	0	0	0
21	19809	1	158	78.0	110	70	1	1	0	0	1	0
23	14532	2	181	95.0	130	90	1	1	1	1	1	0
24	16782	2	172	112.0	120	80	1	1	0	0	0	1
25	21296	1	170	75.0	130	70	1	1	0	0	0	0
27	16747	1	158	52.0	110	70	1	3	0	0	1	0
28	17482	1	154	68.0	100	70	1	1	0	0	0	0

Таблица 2

Имя поле	Описание
id	Идентификатор пациента
age	Возраст в днях
gender	Пол (демографический)
height	Высота в см
weight	Вес в кг
ap_hi	Артериальное давление (высокое)
ap_lo	Артериальное давление (низкое)
cholesterol	Холестерин (обследование)
gluc	Глюкоза (обследование)
smoke	Курение (социальная история)
alco	Употребление алкоголя
active	Физическая активность (социальная история)
cardio	Есть кардиологические проблемы или нет

Каждый x_h может принимать значения из V_h множества: $V_h = \{v_{h1}, v_{h2}, \dots\}$.

Например, переменная «smoke» может принимать значения $\{0,1\}$, который 0 означает не курит, 1 курит. А то есть «cardio» может принимать значения $\{0,1\}$ который 0 означает нету кардиологической проблемы, 1 есть.

Предположим, у нас есть пациент, которого нужно прогнозировать если у него кардиологической проблемы при следующих значений (таблица 3):

Для определения склонности пациента к сердечно-сосудистыми заболеваниями используем классификатор Наивного Байеса по следующей формуле:

$$P(y = v_r | E) = \frac{P(E | y = v_r) \cdot P(y = v_r)}{P(E)}. \quad (1)$$

Таблица 3

переменные	параметры	значение
x_1	age	17474
x_2	gender	1
x_3	height	156
x_4	weight	56.0
x_5	ap_hi	100
x_6	ap_lo	60
x_7	cholesterol	1
x_8	gluc	1
x_9	smoke	0
x_{10}	alco	0

где $P(y = v_r)$ - вероятность того, o_j относится к классу v_r (в нашем случае cardio). Событие, соответствующее равенству независимых переменных определенным значениям, обозначим как E , а вероятность его наступления $P(E)$.

Воспользуемся языком программирования python и библиотекой sklearn. Сперва загружаем набор данных из адреса[12], который состоит из 77 тыс строк и 13 столбцов.

```
import pandas
data = pandas.read_csv("cardio.csv", sep=';')
data.head()
```

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0

Для независимых переменных X присваиваем значений столбцов диапазоне [0,13) а для [13].

```
import numpy as np
X = data[data.columns[0:-1]]
Y = data[data.columns[-1]]

from sklearn.naive_bayes import GaussianNB
clf = GaussianNB()
clf.fit(X, Y)
```

Далее для входного параметра даём параметры пациента.

```
pred = V[17474,1,156,56.0,100,60,1,1,0,0]
result =clf.predict(pred)
print(result)

[0]
```

Результат работы алгоритма показывает, что у пациента, в настоящее время, нет рисков развития сердечно-сосудистых заболеваний.

4 Заключение

Показатели сердечно-сосудистых заболеваний являются серьезной проблемой при анализе медицинских данных, поскольку они стали одной из основных причин смертности. Машинное обучение может улучшить понимание врачей, особенно в прогнозировании сердечных заболеваний, что позволит им лучше адаптироваться к диагностике и лечению пациентов. Цель задачи статьи была продемонстрировать возможность алгоритмов машинного обучения. Качество эксперимента была выбрано алгоритм Наивного Байеса.

Литература

- [1] *Krittanawong C., Virk H.U.H., Bangalore S., et al.* Machine learning prediction in cardiovascular diseases: a meta-analysis. // Scientific Reports. – Vol. 10. No 1. – 2020. – P. 16057–16111.
- [2] *Duan W., Xu C., Liu Q. et al.* Levels of a mixture of heavy metals in blood and urine and all-cause, cardiovascular disease and cancer mortality: a population-based cohort study. // Environmental Pollution. – Vol. 263. – 2020.
- [3] *Han D., Kolli K.K., Al'Aref S.J. et al.* Machine learning framework to identify individuals at risk of rapid progression of coronary atherosclerosis: from the PARADIGM registry. // Journal of American Heart Association. – 2020. – Vol. 9. No. 5.
- [4] *Krittanawong C., Virk H.U.H., Bangalore S. et al.* Machine learning prediction in cardiovascular diseases: a meta-analysis. // Sci Rep 10. – 2020. doi: <http://dx.doi.org/10.1038/s41598-020-72685-1>
- [5] *Islam M.R., Osmani M.B., Tuhin S.K.* IoT-based Low Cost Healthcare System for Cardiac Patient of Bangladesh using Bio-sensors. // 2nd International Conference on Innovation in Engineering and Technology (ICIET). – 2019. – P. 1–6.
- [6] *Sajja T.K., Kalluri H.K.* A deep learning method for prediction of cardiovascular disease using convolutional neural network. // Revue d'Intelligence Artificielle. – Vol. 34. No. 5 – 2020. – P. 601–606. doi: <http://dx.doi.org/10.18280/ria.340510>
- [7] *Alqahtani A., Alsubai S., Sha M., Vilcekova L., Javed T.* Cardiovascular Disease Detection using Ensemble Learning. // Comput Intell Neurosci. – 2022. doi: <http://dx.doi.org/10.1155/2022/5267498>
- [8] *Kumar R., Kumar P., Tripathi R., Gupta G. P., Islam A.K., Shorfuzzaman M.* Permissioned blockchain and deep-learning for secure and efficient data sharing in industrial healthcare systems. // IEEE Transactions on Industrial Informatics. – 2022. – P. 1. doi: <http://dx.doi.org/10.1109/tii.2022.3161631>
- [9] *Mohan S., Thirumalai C., Srivastava G.* Effective heart disease prediction using hybrid machine learning techniques. // IEEE Access. – 2019. doi: <http://dx.doi.org/10.1109/access.2019.2923707>

- [10] *Au H., Li J.P., Memon M.H., Nazir S., Sun R.* A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. // *Mobile Information Systems*. – 2018. doi: <http://dx.doi.org/10.1155/2018/3860146>.3860146
- [11] *Renugadevi G., Asha Priya G., Sankari B.D., Gowthamani R.* Predicting heart disease using hybrid machine learning model. // *Journal of Physics: Conference Series*. – 2021.
- [12] <https://github.com/iryndin/Machine-Learning/tree/master/data/01-cardio>

Поступила в редакцию 13.01.2023

UDC 004.4

MACHINE LEARNING BASED PREDICTION OF CARDIOVASCULAR DISEASES

^{1*}*Ravshanov N.,* ¹*Pekos O.A.,* ²*Bakaev I.I.*

*ravshanzade-09@mail.ru

¹Digital Technologies and Artificial Intelligence

Development Research Institute,

100124, 17A Buz-2, Mirzo-Ulugbek district, Tashkent, Uzbekistan;

²Bukhara State University,

705018, Muhammad Ikbol 11, Bukhara, Uzbekistan.

Since the second half of the last century, cardiovascular diseases have firmly taken the top lines in the structure of mortality worldwide, and in some countries - more than two thirds of the total number of deaths annually. In this regard, an extremely urgent problem is their early diagnosis, which greatly increases the success of treatment and the likelihood of a positive outcome of the disease. Conventional CVD risk assessment methods implicitly allow for a linear relationship between risk factors and CVD outcome. In addition, CVD risk is usually calculated using only clinical information without taking into account other important factors such as social determinants of health. Since machine learning has long been a generally accepted means of detecting patterns in data, this class of artificial intelligence methods is actively used in clinical cardiology today. Including when solving problems: automatic recognition and decoding of digital signals of ECG, EchoCG, XM ECG; diagnostics of pathologies; automatic classification of patients according to various criteria; assessing the risks of developing adverse events in the cardiovascular system; automatic selection of treatment tactics; predicting the results of cardiac rehabilitation, etc. The purpose of this work was to solve the problem of assessing the propensity of patients to cardiovascular diseases by applying one of the popular machine learning methods, the naive Bayes classifier, to a dataset of outpatient examinations of patients. The algorithm was implemented in python using the scikit-learn library.

Keywords: cardiovascular disease, machine learning algorithms, Naive Bayes classifier.

Citation: Ravshanov N., Pekos O.A., Bakaev I.I. 2023. Machine learning based prediction of cardiovascular diseases. *Problems of Computational and Applied Mathematics*. 1(46): 109-115.