

MUHAMMAD AL-XORAZMIY
AVLODLARI
ILMIY-AMALIY VA AXBOROT-
TAHLILY JURNAL

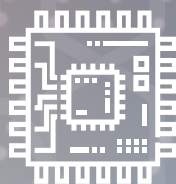
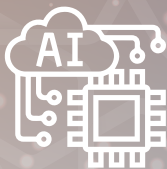
DESCENDANTS OF MUHAMMAD
AL-KHWARIZMI
SCIENTIFIC-PRACTICAL AND
INFORMATION-ANALYTICAL JOURNAL



1(31)/2025

ISSN-2181-9211

MUHAMMAD AL-XORAZMIY NOMIDAGI
TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI



MUHAMMAD AL-XORAZMIY AVLODLARI

Ilmiy-amaliy va axborot-tahliliy jurnal 2017 yilda
ta'sis etilgan

1(31)/2025

Tahririyat kengashi a'zolari

Maxkamov B.SH. – Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti (TATU) rektori, Tahririyat kengashi raisi

Sultanov Dj.B.. – Tahririyat kengashi raisi o'rinbosari

Tashev K.A. – Tahrir kengashi raisi o'rinbosari

Nosirov X.X. – DSc., dots. bosh muharrir

Raximov B.N. – t.f.d., prof. bosh muharrir o'rinbosari

Muharrirlar:

Kamilov M.M. – t.f.d., prof., akademik.

Musayev M.M. – t.f.d., prof.

Abduraxmonov K.P. – f.-m.f.d., prof.

Jumanov J.X. – t.f.d., prof.

Muxamediyeva D.T. – t.f.d., prof.

Isayev R.I. – t.f.n., prof.

Yusupov A. – f.-m.f.d., prof.

Yakubova M.Z. – t.f.d., prof. (Qozog'iston)

Xalikov A.A. – t.f.d., prof. (TDTU)

Nazarov A.M. – t.f.d., prof. (TDTU)

Jmud V.A. – professor (Rossiya)

Miroslav Skoric – professor (Avstriya)

Dzhurakhalov A. – professor (Belgiya)

Abrarov S.M. – professor (Kanada)

Kyamakya K. – professor (Avstriya)

Chedjou J.Ch. – professor (Avstriya)

Davronbekov D.A. – t.f.d., prof.

Anarova Sh.A. – t.f.d., prof.

Pisetskiy Y.V. – t.f.d., prof.

Nishonov A.X. – t.f.d., dots.

Muminov B.B. – t.f.d., prof.

Khudayberdiyev M.X. – t.f.d., prof.

Raximov N.O. – t.f.d., dots.

Amirsaidov U.B. – t.f.d., dots.

Kerimov K.F. – t.f.d., dots.

Ganiyev A.A. – t.f.n., dots.

Gavrilov I.A. – t.f.n., dots.

Gubenko V.A. – t.f.n., dots.

Pulatov Sh.U. – t.f.n., dots.

Kutlimuratov A. – PhD, dots.

Shaxobiddinov A.SH. – PhD, dots.

Madaminov X.X. – PhD, dots.

Xudaybergenov T.A. – PhD, dots.

Ro'ziboyev O.B. – PhD, dots.

Yaxshibayev D.S. – PhD, dots.

Mirsagdiyev O.A. – PhD, dots.

Puziy A.N. – PhD, dots.

Saymanov I.M. – PhD, dots.

Aripova U.X. – PhD, dots.

Berdiyev A.A. – PhD, bosh muharrir yordamchisi

Xudayberganov J.D. – texnik muxarrir

Kengesbayev S.K. – texnik muxarrir

MUNDARIJA

DASTURIY VA KOMPYUTER INJINIRING TEXNOLOGIYALARINING ZAMONAVIY MUAMMOLARI

Камилов М.М., Тожибоев Б.М., Равшанов А.А. Идентификация опорных множеств признаков на изображениях дистанционного зондирования земли	3
Мухамедиева Д.Т., Раупова М.Х. Решение задач диагностики диабета с применением машинного обучения	10
Рaxmatov F.A., Abdirazakov F.B., Temirov A.A., Nasirov S.U. Nutq signallari yordamida shaxsni tanib olishda sun'iy intellekt asosidagi samarali algoritmlarini qo'llash	17
Zaynidinov H.N., Mamirov X.X., Yusupov I., Boytemirov A.M. An optimal algorithm for monitoring heart activity based on IoT and cloud technologies	23
Ismailov M.A., Avezov T.A. Elastomer kompozitlar ishlab chiqarsihda ko'p qatlamli polimer mahsulotlar uchun optimal vulkanizatsiya vaqtini hisoblash algoritmi	28
Zaynidinov H.N., Mahmudjanov S.U., Aliqulov A.X. Tashxislashga ko'maklashuvchi axborot tizimida tashxis qo'yilgan bemorlar to'g'risidagi ma'lumotlarni optimallashtirish usul va algoritmlari	31
Axatov A.R., Ximmatov I.Q. Shaxsni ob'yekt sifatida aniqlash uchun tasvirdan tashqi ko'rinishni modellashtirish	41
Asqaraliyev O.U. Ma'lumotlarni ishlashning yagona kompyuter tizim markaziy server hisoblash klasterlaridagi yuklamalarni muvozanatlash jarayonining modeli va protseduralari	48
Sultanov Dj.B., Axmedova X.X. O'zbek tilida gaplarni perefraz qilish masalasi	53
Meliqo'ziyev R.Sh., Meliqo'ziyev M.Sh., Sabirova U.Sh. Fourier transformasi va statistik modellar asosida multimedia ma'lumotlarining ruxsatsiz o'zgartirilishini aniqlash	59
Allamov O.T., Babajanov B.F. Dasturiy tizimlar o'rtaasida ma'lumotlarni o'zaro muvofiqlikda ishlashini tashkil qilishni turli usullardagi tahlili	63
Nishanov A.X., Xaydarov Sh.I., Mamajanov R.Ya., Yuldashev R.R. Su bezi saratoni kasalligini erta tashxislashga qaratilgan sintetik o'quv tanlanmani yaratish algoritmi	67
OPTIK ALOQA TIZIMLARI, TELEKOMMUNIKATSIYA TARMOQLARI VA KOMMUTATSIYA TIZIMLARINING RIVOJLANISH TAMOYILLARI	
Parsiye S.S., Badalov J.I., Raxmatov I. Yu. Telekommuniukatsiya tarmoqlarida aloqa tugunlarini optimal joylashtirish	75
Eshniyazova G. A Privacy Preserving Framework for Smart Cities based on IoT, Smart Buildings and Big Data	80
Kengesbayev S.K. Suyuq muhitni kuzatish uchun optoelektron konvertorning matematik modeli	87
Karimova A.O. Analysis of MIMO wireless communication technology	92
Gulomov Sh.R., Tashev S.N. Tarmoq xavfsizligi uchun paketli filtrlash texnikasi	96
Бадалов Ж.И., Давлетова Х.Р. Анализ технологий оптической сети доступа	99
Khayrullaev A.F. Modeling of environmental monitoring process using fiber optic sensors	103
Shamsiyev A.S. Simsiz sensor tarmoqlar uchun takomillashtirilgan xavfsiz va energiya tejankor marshrutlash protokolini loyihalashtirish	108
Saitkamolova G.K. Telekommunikatsiya tarmog'ida trafik xususiyatlarini baholash masalasi	112
RAQAMLI TELEVIDENIYE VA RADIOESHITTIRISH, SIMSIZ TEXNOLOGIYALAR VA RADIOTEKNIKANI RIVOJLANTIRISH ISTIQBOLLARI	
Нуруллаев Е.Е. Эффективное разделение носителей заряда в фотодетекторах на основе гетеропереходов	120
Abdulkhayev N.M., Khudayberganov J.D. Transmission and processing	126

Muassis:

Muhammad al-Xorazmiy nomidagi
Toshkent axborot texnologiyalari
universiteti

Manzil:

100084, O'zbekiston, Toshkent sh., Amir
Temur ko'chasi, 108
Telefon: 71 238-64-38;
e-mail: alxorazmiy@tuit.uz
Jurnal sayti: <http://alxorazmiy.uz>

Bosishga ruxsat etildi:

Qog'oz bichimi 60x84 1/8
Bosma tabog'i 15,5. Adadi 100 nusxa
Buyurtma raqami №195 "Fan va
texnologiyalar Markazining
bosmaxonasi" da chop etildi
Toshkent shahri Olmazor ko'chasi, 171.
Jurnal O'zbekiston Matbuot va
axborot agentligida 2017 yil
22 iyunda 0921 raqami bilan ro'yxatdan
o'tgan.
Jurnal yilda 4 marotaba
(har chorakda) chop etiladi.

of real-time signals using the wifi NOTEMCU 8266 CH340 module in a virtual field created in a virtual reality simulator	
Абидова Н.Б. Способы обнаружения и противодействия с малогабаритными беспилотными летательными аппаратами	130
Губенко В.А., Арипова У.Х., Бердиев А.А., Алимухамедова М.Э., Кан В.С. Исследования влияния механических воздействий на характеристики коаксиальных волноводов	134
Mamatov N.S., Sultanov J.B., K.K.Erejevov., Narzullayev I.S., Samijonov B.N. Yuz tasvirlarini tanib olishning lokal binar shablonlarga asoslangan algoritmi	140
Gaziyev X., Berdiyev A., Vohobov M., Kurbanov Sh. Takomillashtirilgan Brayl texnologiyalari: muhandislik yondashuvlari va funksional qiyoslash	146
Djalilov B.O. Xarakteristik modalar nazariyasidan foydalangan holda uyali aloqa qurilmalari antenna parametrlarini optimallashtirish	150
Donaboyev Sh.Sh., Abdullayev R.B. Comparative Analysis of Reliable Discrete Devices Based on FPGAs	154
Писецкий Ю.В., Вотинков К.А., Дадамухамедова Д.Д. Исследование параметров оценки эффективности передачи сигналов в системах спутниковой связи	161
Aliyev U.T., Madaminov H.X., Murodov A.Dj. 5G tarmoqlarida radiochastotalar resurslarini dinamik boshqarishni tashkil etishning o'ziga xos xususiyatlari	165
Yusupov Ya.T. Amplitudasi modulyatsiyalangan signallarni demodulyatsiyalash jarayonlarini o'rgatishda NI multismidan foydalanish	175
Safarov A.M., Sattarov Kh.A. The principle of construction of direct current converters and their implementation	180
Bayjanova I.E., Sabirova U.Sh. Chiziqli model va yopiq kontur usuli asosida kenglik-impulsi modulyatsiyalashli asosiy radiotexnik qurilmalarning impedanslarini qiyosiy tahlil qilish	184
Hotamov A., Nurmurodov J.X. Simsiz tizimlarida ma'lumotlar oqimlarini boshqarish metodologiyasi	190
ELEKTRON HUKUMAT	
Gafurov Sh.R. Elektron hukumat tizimida axborot xavfsizligining buzilishini aniqlash usuli va algoritmi	194
Жолмурзаев Т.А. Цифровая трансформация в государственном управлении Узбекистана	200
ILMIY AXBOROTLAR	
Igamberdiyev K.A. Mathematical Modeling of the Lyapunov Function of a Nonlinear Dynamic System	203
Abdulxayev N.M. Oliy ta'lim muassalarida tahsil olayotgan talabalarni virtual reallikka asoslangan ko'zoynak orqali o'qitish samaradorligini baholash	206
Baltayev R.Sh., Ishmetov B.Y., Yusupova Sh.B., Nurmetova B.B., Abdullayev A.Sh., Ataxanov M.F. Creating a telegram bot using the aiogram package of the python programming language	209
Amurova N.Y. Modern Developments and Trends in the Field of Green Technologies in the Republic of Uzbekistan	214
Сейтназаров К.К., Юсупова М.Т., Туремуратова Б.К. Анализ деградации растительности в приаралье с использованием аэрокосмических данных и спектральных индексов	223
Ruzmetov A.A. Optimization of algorithms for interacting ros2 and hardware interfaces	229
Sadullayev J.Sh., Atabayev S.S., Xudayberganov T.A. 2D geometric shapes detection using OpenCV	234
Жураева К.К., Калибаева А.М. Анализ состояния и перспективы развития солнечной энергетики в Узбекистане	238
Maxmudjanov S.U., Mamatov A.Sh., Maxanov S.M. Yozma ishlarni baholashda mutaxassis va sun'iy intellekt samaradorligining qiyosiy tahlili	242
Abdullayev E.A. Shamol qurilmasiga ega bo'lgan korxonalarining yuklama grafiklarini optimallashtirish	250
Абидов Б.К., Джабборов М.А. Методы оценки качества услуг систем телерадиовещания	254
Uteuliev N.U., Djaykov G.M., Sagidullaev N.I. Tibbiyot muassalarini avtomatlashtirish: yondashuvlar va matematik modellashtirish	258
Анарова Ш.А., Яхшибаев Д.С. Исследование геометрически нелинейных задач изгиба термоупругих пластин со сложной конфигурацией	261
Яхшибоев Д.С. Моделирование движение стратификационного течения в водохранилищах	271
Xujamatova Sh. Telekommunikatsiya sohasida kontentni yetkazib berish tarmog'ini joriy etishning iqtisodiy samaradorlik mezon va ko'rsatkichlari	275

Мухамедиева Д.Т., Раупова М.Х.

Решение задач диагностики диабета с применением машинного обучения

Диагностика диабета является важной задачей в медицинской практике, учитывая глобальный рост числа пациентов с этим хроническим заболеванием. Машинное обучение, с его способностью анализировать и извлекать паттерны из больших объемов данных, предоставляет мощный инструмент для улучшения процессов диагностики и управления диабетом. В этом контексте, машинное обучение применяется в различных аспектах диагностики диабета, начиная с ранней диагностики и заканчивая индивидуализированным лечением и мониторингом. Это открывает новые возможности для более эффективного предотвращения и управления этим заболеванием. В данной работе мы рассмотрим разнообразные применения машинного обучения в задачах, связанных с диагностикой диабета, а также осветим роль машинного обучения в усилении современной медицинской практики и улучшении качества жизни пациентов, страдающих диабетом.

Ключевые слова: диабет, машинное обучение, модель, Датасет Diabetes, классификация.

1. Введение.

Диабет — это хроническое заболевание, которое влияет на миллионы людей по всему миру. Для ранней диагностики и эффективного управления диабетом медицинская община все больше обращается к методам машинного обучения и анализу данных. Актуальность машинного обучения в контексте диабета нельзя недооценивать. Это актуальное исследовательское и практическое направление, которое имеет ряд важных аспектов. Диабет становится все более распространенным заболеванием, и его эпидемия неуклонно растет. Машинное обучение предоставляет инструменты для более эффективной диагностики и управления заболеванием. Ранняя диагностика диабета и его типа является ключевым фактором для успешного лечения и предотвращения осложнений. Машинное обучение позволяет разрабатывать модели для ранней диагностики. Каждый пациент уникален, и машинное обучение может помочь создавать индивидуальные планы лечения, учитывая генетические, клинические и лабораторные данные. С медицинскими данными сталкиваются разные источники и форматы. Машинное обучение позволяет интегрировать и анализировать эти данные, создавая более полное представление о состоянии пациентов. Осложнения диабета могут быть опасными. Модели машинного обучения могут предсказывать вероятность их развития, что позволяет предпринимать меры для их предотвращения [1-4].

Датасет Diabetes является ценным ресурсом для исследователей и практиков в области медицинского анализа данных и машинного обучения. Он позволяет разрабатывать и тестировать модели для предсказания и мониторинга прогрессии диабета и может способствовать улучшению подходов к диагностике и управлению этим серьезным хроническим заболеванием. Набор данных Diabetes включает 10 числовых признаков, которые описывают состояние пациентов. Эти признаки представляют собой важные медицинские измерения и характеристики здоровья. Вот более подробное описание каждого из этих признаков [5-6]:

Уровень глюкозы в крови (Blood Sugar Level): Этот признак представляет собой уровень глюкозы в

крови пациента. Это важное измерение для диагностики диабета и мониторинга его контроля.

Артериальное давление (Blood Pressure): Этот признак представляет собой артериальное давление пациента. Высокое артериальное давление может быть связано с сердечно-сосудистыми заболеваниями, которые часто сопровождают диабет.

Индекс массы тела (BMI - Body Mass Index): BMI измеряет отношение веса к росту пациента и используется для определения избыточного веса или ожирения, что является фактором риска для развития диабета.

Индекс уровня молекулы сердечного заболевания (Serum Levels of Heart Disease Marker): Этот признак представляет собой индекс уровня определенной молекулы, который может быть связан с риском развития сердечно-сосудистых заболеваний у пациентов с диабетом.

Уровень серума липопротеина низкой плотности (Low-Density Lipoprotein Cholesterol): Этот признак измеряет уровень "плохого" холестерина в крови, который также может быть связан с сердечно-сосудистыми рисками.

Уровень серума липопротеина высокой плотности (High-Density Lipoprotein Cholesterol): Этот признак измеряет уровень "хорошего" холестерина в крови и является важным фактором для оценки сердечно-сосудистого здоровья.

Логарифм натуральный трех инсулиновых параметров (Natural Log of Three Insulin Parameters): Этот признак представляет собой натуральный логарифм от трех параметров, связанных с инсулином и обменом веществ.

Возраст (Age): Этот признак представляет собой возраст пациента на момент наблюдения. Возраст может быть важным фактором в развитии и прогрессии диабета.

Пол (Sex): Этот признак отражает пол пациента (мужской или женский) и может иметь влияние на риск развития диабета.

Уровень С-пептида (C-Peptide Level): С-пептид является молекулой, производимой вместе с инсулином. Его уровень может быть связан с функцией поджелудочной железы и уровнем инсулина.

Эти числовые признаки предоставляют важную информацию о состоянии пациентов и могут использоваться для разработки моделей машинного обучения, предсказывающих прогрессию диабета и оценивающих риски [7].

Цель применения машинного обучения в контексте диабета может варьироваться в зависимости от конкретной задачи и сценария, но общие цели включают в себя ранняя диагностика и прогнозирование риска. Целью является создание моделей, которые способны рано диагностировать диабет или предсказывать риск его развития. Это позволяет начать лечение на ранних стадиях заболевания и предотвратить его осложнения [8].

Машинное обучение может помочь разрабатывать индивидуальные планы лечения и мониторинга для каждого пациента с учетом его уникальных характеристик и реакции на лечение. Машинное обучение может быть направлено на мониторинг данных пациентов с целью предупреждения осложнений, таких как гипогликемия или диабетическая кетоацидоз. Главной целью машинного обучения в контексте диабета является обеспечение более высокого качества ухода для пациентов, снижение риска осложнений и улучшение их качества жизни [9].

Постановка задачи машинного обучения в контексте диабета зависит от конкретных целей и сценариев [10].

Классификация типа диабета: Задачей может быть классификация типа диабета на основе клинических и лабораторных данных пациентов. Тип 1 и тип 2 диабет имеют разные характеристики и требуют разных подходов к лечению.

Диагностика диабета: Создание модели для диагностики диабета у лиц, подозреваемых в наличии заболевания. Это может включать в себя анализ уровня глюкозы в крови, антропометрических данных и других признаков.

Прогнозирование риска развития диабета: Задачей может быть прогнозирование риска развития диабета у лиц, не имеющих заболевания в данный момент. Это помогает рано выявить пациентов с повышенным риском и предложить им меры профилактики.

Прогнозирование прогрессии диабета: Создание моделей для прогнозирования темпов прогрессии диабета у пациентов, уже страдающих заболеванием. Это может помочь врачам оптимизировать лечение и мониторинг.

Оптимизация лечения: Задачей может быть разработка моделей, которые помогают врачам выбирать наилучшие методы лечения для каждого пациента, учитывая его характеристики и реакцию на лекарства.

Мониторинг и предупреждение осложнений: Задачей может быть создание систем мониторинга, которые предупреждают осложнения, такие как гипогликемия, диабетическая кетоацидоз или диабетическая ретинопатия.

Интеграция медицинских данных: Задачей может быть интеграция данных из различных источников, чтобы предоставить врачам более полное представление о состоянии пациентов и поддержать принятие решений.

Исследование новых методов лечения: Задачей может быть анализ данных клинических исследований для разработки новых методов лечения диабета.

Создание систем поддержки принятия решений: Модели машинного обучения могут служить в качестве систем поддержки принятия решений для врачей и пациентов, предоставляя информацию о рисках и возможных вариантах лечения.

Оценка эффективности лечения: Задачей может быть оценка того, насколько эффективно лечение действует на конкретного пациента, и корректировка плана лечения, если необходимо.

Задачи машинного обучения в области диабета способствуют более эффективной диагностике, лечению и уходу за пациентами, а также содействуют научным исследованиям и разработке новых методов борьбы с этим хроническим заболеванием.

2. Методы.

Для решения задач в области диагностики, управления и исследования диабета могут использоваться различные методы машинного обучения. Вот некоторые из наиболее распространенных методов [10-11]:

1. Линейная регрессия может быть использована для задачи прогнозирования, например, прогнозирования уровня глюкозы в крови на основе различных параметров.

Линейная регрессия — это метод для предсказания значения зависимой переменной на основе линейной комбинации одной или нескольких независимых переменных. Вот алгоритм для обучения модели линейной регрессии:

Загрузка данных:

Сначала необходимо загрузить набор данных, содержащий значения независимых (факторных) и зависимых переменных. Обычно данные представляются в форме таблицы или массива.

Подготовка данных:

Проверьте данные на отсутствие нулевых значений и выбросов.

Разделите данные на обучающий набор и тестовый набор, чтобы оценить производительность модели.

Определение модели:

Линейная регрессия моделирует зависимую переменную (Y) как линейную комбинацию независимых переменных (X) с весами (коэффициентами) и константным членом (смещением).

Модель может быть представлена уравнением: $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$, где b_0 — это смещение (intercept), а b_1, b_2, \dots, b_n — коэффициенты.

Обучение модели:

Для обучения модели линейной регрессии используется метод наименьших квадратов (Least Squares). Этот метод находит значения коэффициентов, которые минимизируют сумму квадратов разностей между предсказанными и фактическими значениями зависимой переменной.

Можно использовать библиотеки Python, такие как scikit-learn, для обучения модели линейной регрессии.

Прогнозирование:

После обучения модели вы можете использовать ее для прогнозирования значений зависимой переменной на новых данных.

Оценка модели:

Оцените производительность модели на тестовом наборе данных, используя метрики, такие как средняя квадратичная ошибка (MSE) и коэффициент детерминации (R-squared).

Интерпретация результатов:

Проанализируйте коэффициенты модели, чтобы определить, какие независимые переменные оказывают наибольшее влияние на зависимую переменную.

Развертывание модели:

Если модель удовлетворяет ваши требования, вы можете развернуть ее для использования в реальных приложениях.

Это общий алгоритм для линейной регрессии. Реализация конкретных шагов может различаться в зависимости от используемого программного обеспечения и языка программирования.

2. Модель наивного байесовского классификатора [1].

Загружаем данные Diabetes и преобразуем задачу в бинарную классификацию на основе медианного значения.

Разделяем данные на обучающий и тестовый наборы, а также стандартизируем признаки (наивный байес не требует стандартизации, но мы это сделали для согласованности с предыдущими примерами).

Создаем и обучаем модель наивного байесовского классификатора.

Делаем прогноз на тестовом наборе данных и рассчитываем метрики.

Выводим матрицу путаницы, отчет о классификации и график AUC-ROC.

3. Метод опорных векторов (Support Vector Machines, SVM) [2] SVM может применяться как для задач классификации, так и для регрессии. Он особенно полезен в случаях, когда данные разделяются нелинейно.

Алгоритм Support Vector Machine (SVM) - это метод машинного обучения, используемый для задач классификации и регрессии. Вот общий алгоритм SVM для задачи бинарной классификации:

Загрузка данных:

Сначала загрузите набор данных, содержащий набор объектов (векторов признаков) и соответствующие им метки классов. В задаче

бинарной классификации есть два класса: положительный (1) и отрицательный (0).

Подготовка данных:

Проверьте данные на отсутствие нулевых значений и проведите предварительную обработку, если это необходимо. Масштабируйте признаки, чтобы они имели одинаковый диапазон.

Выбор ядра (Kernel):

SVM может использовать разные ядра, такие как линейное, полиномиальное или радиально-базисное функциональное ядро (RBF). Выберите подходящее ядро для вашей задачи.

Обучение модели:

Обучите SVM на обучающем наборе данных, используя выбранное ядро. Основная цель - найти оптимальную разделяющую гиперплоскость, которая максимизирует отступы (margins) между классами.

Оптимизация параметров:

Настройте параметры SVM, такие как коэффициент регуляризации (C) и параметры ядра, чтобы достичь лучшей производительности модели.

Прогнозирование:

Используйте обученную модель для предсказания меток классов для новых данных.

Оценка модели:

Оцените производительность модели с использованием метрик, таких как точность, полнота, F1-мера и матрица ошибок (confusion matrix) в случае классификации.

Кросс-валидация:

Используйте кросс-валидацию для более надежной оценки производительности модели и избегания переобучения.

Интерпретация результатов:

Анализируйте полученные результаты и веса признаков, чтобы понять, какие признаки вносят наибольший вклад в классификацию.

Регуляризация и настройка:

В зависимости от результатов, вы можете регуляризовать модель или настроить другие параметры, чтобы улучшить производительность.

Развертывание модели:

Если модель удовлетворяет вашим требованиям, вы можете развернуть ее для использования в реальных приложениях.

Это общий алгоритм SVM для задачи бинарной классификации. Для задачи регрессии принципы работы алгоритма похожи, но цель состоит в предсказании непрерывных значений, а не меток классов.

4. Деревья решений и случайные леса [12-14]. Деревья решений могут использоваться для классификации и регрессии. Случайные леса представляют собой ансамбль деревьев решений и могут улучшить предсказательную способность модели.

Загружаем данные Diabetes и преобразуем задачу в бинарную классификацию на основе медианного значения.

Разделяем данные на обучающий и тестовый наборы, а также стандартизируем признаки (стандартизация не обязательна, но мы это сделали для согласованности с предыдущими примерами).

Создаем и обучаем модель дерева решений.

Делаем прогноз на тестовом наборе данных и рассчитываем метрики.

Выводим матрицу путаницы, отчет о классификации и график AUC-ROC.

Это лишь несколько примеров методов машинного обучения, которые могут применяться в контексте диабета. Выбор конкретного метода зависит от характера задачи и доступных данных. Кроме того, часто используются ансамбли методов для улучшения качества прогнозов [15-17].

3. Результаты.

Результаты машинного обучения в области диабета могут быть разнообразными и зависят от конкретной задачи, используемых методов и доступных данных. Вот некоторые из типичных результатов, которые можно достичь с помощью машинного обучения в контексте диабета:

1. Для решения задачи регрессии на наборе данных Diabetes из библиотеки scikit-learn в Python, вы можете использовать следующий алгоритм:

Mean Squared Error: 2900.1732878832318

R-squared: 0.452606602161738

Среднеквадратичная ошибка (Mean Squared Error, MSE) и коэффициент детерминации (R-squared) - это две основные метрики, используемые для оценки производительности модели линейной регрессии. Ваши значения MSE и R-squared говорят о том, насколько хорошо ваша модель соответствует данным и предсказывает зависимую переменную.

Среднеквадратичная ошибка (MSE):

MSE измеряет среднеквадратичное отклонение между фактическими значениями зависимой переменной (в вашем случае, вероятно, уровень глюкозы в крови) и предсказанными значениями, вычисленными вашей моделью.

Ваше значение MSE равно 2900.17. Чем меньше MSE, тем лучше. Это означает, что ваша модель в среднем ошибается на квадратный корень из этой величины при прогнозировании значений.

Коэффициент детерминации (R-squared):

R-squared измеряет, насколько хорошо ваша модель объясняет вариабельность зависимой переменной. Он находится в диапазоне от 0 до 1, где 0 означает, что модель не объясняет вариабельность, а 1 означает, что модель идеально соответствует данным.

Значение R-squared равно 0.4526. Это означает, что ваша модель объясняет примерно 45% вариабельности в данных. То есть модель объясняет менее половины изменений в зависимой переменной, что может быть улучшено с помощью более сложных моделей или дополнительных признаков.

Итак, модель имеет среднее качество предсказания (MSE не слишком высокий), но она объясняет только часть вариабельности в данных (R-

squared не очень близок к 1). Вам может потребоваться дополнительно анализировать данные, добавить новые признаки или выбрать другой метод моделирования, чтобы улучшить производительность модели.

2. Модель наивного байесовского классификатора

Confusion Matrix:

[[37 12]

[13 27]]

Элемент (0,0) (верхний левый угол) представляет собой количество истинно отрицательных (TN) примеров.

Элемент (0,1) (верхний правый угол) представляет собой количество ложноположительных (FP) примеров.

Элемент (1,0) (нижний левый угол) представляет собой количество ложноотрицательных (FN) примеров.

Элемент (1,1) (нижний правый угол) представляет собой количество истинноположительных (TP) примеров.

Classification Report:

	precision	recall	f1-score	support
Class 0	0.74	0.76	0.75	49
Class 1	0.69	0.68	0.68	40
accuracy		0.72		89
macro avg	0.72	0.72	0.72	89
weighted avg	0.72	0.72	0.72	89

Accuracy: 0.7191

Precision: 0.6923

Recall: 0.6750

F1-Score: 0.6835

AUC-ROC: 0.8260

Classification Report предоставляет детальные метрики для каждого класса (Class 0 и Class 1), а также средние значения (macro avg и weighted avg). В вашем случае:

Precision (точность) измеряет, как много из объектов, которые модель предсказала как положительные, действительно являются положительными. Precision для Class 0 составляет 0.74, а для Class 1 - 0.69.

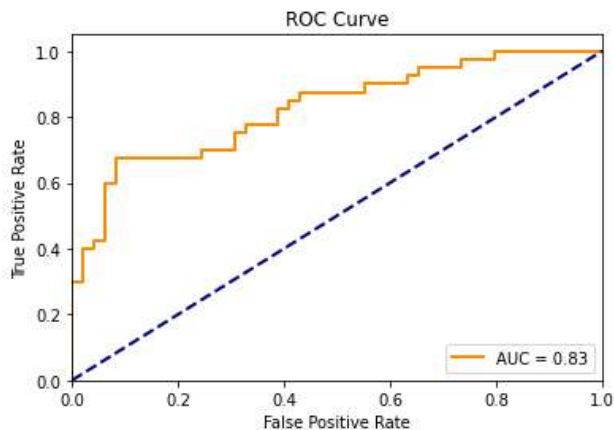
Recall (полнота) измеряет, как много из всех действительных положительных объектов модель правильно классифицировала. Recall для Class 0 составляет 0.76, а для Class 1 - 0.68.

F1-Score - это гармоническое среднее между точностью и полнотой. F1-Score для Class 0 составляет 0.75, а для Class 1 - 0.68.

Accuracy (точность) - доля верно классифицированных примеров относительно общего числа примеров. Accuracy составляет 0.7191, что означает, что модель правильно классифицировала 71.91% всех примеров.

AUC-ROC (Площадь под ROC-кривой):

AUC-ROC измеряет площадь под кривой ROC (Receiver Operating Characteristic), которая отображает производительность модели при различных порогах классификации. Значение AUC-ROC близкое к 1 (ваше значение 0.8260) указывает на хорошую способность модели разделять классы.



Эти метрики позволяют оценить производительность модели классификации, в частности, способность модели разделять классы, сбалансированность точности и полноты, а также общую точность модели. В вашем случае, модель показала средние результаты в оценке классификации.

3.SVM с линейным ядром.

Загружает данные Diabetes и преобразует задачу в бинарную классификацию (на основе медианного значения).

Разделяет данные на обучающий и тестовый наборы, а также стандартизирует признаки.

Создает и обучает модель SVM с линейным ядром.

Делает прогноз на тестовом наборе данных и рассчитывает метрики путаницы, accuracy, precision, recall, F1-score и AUC-ROC.

Строит график ROC-кривой для визуализации производительности модели.

Confusion Matrix:

[[35 14]

[10 30]]

Classification Report:

	precision	recall	f1-score	support
Class 0	0.78	0.71	0.74	49
Class 1	0.68	0.75	0.71	40

accuracy		0.73	89	
macro avg	0.73	0.73	0.73	89
weighted avg	0.73	0.73	0.73	89

Accuracy: 0.7303

Precision: 0.6818

Recall: 0.7500

F1-Score: 0.7143

AUC-ROC: 0.8398

Матрица путаницы показывает количество верно и ошибочно классифицированных примеров для каждого класса.

Classification Report (Отчет о классификации):

Classification Report предоставляет детальные метрики для каждого класса (Class 0 и Class 1), а также средние значения (macro avg и weighted avg). В вашем случае:

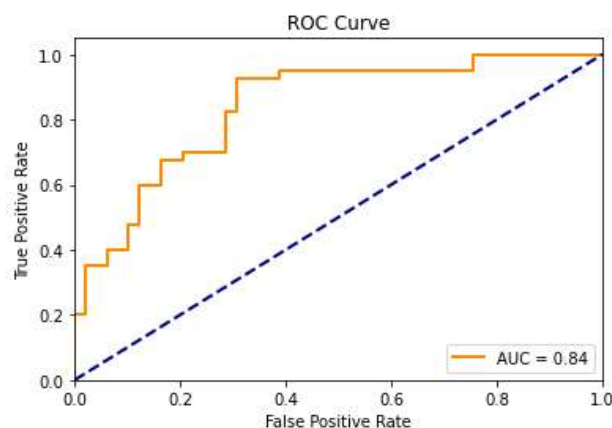
Precision (точность) измеряет, как много из объектов, которые модель предсказала как положительные, действительно являются положительными. Precision для Class 0 составляет 0.78, а для Class 1 - 0.68.

Recall (полнота) измеряет, как много из всех действительных положительных объектов модель правильно классифицировала. Recall для Class 0 составляет 0.71, а для Class 1 - 0.75.

F1-Score - это гармоническое среднее между точностью и полнотой. F1-Score для Class 0 составляет 0.74, а для Class 1 - 0.71.

Accuracy (точность) - доля верно классифицированных примеров относительно общего числа примеров. Accuracy составляет 0.7303, что означает, что модель правильно классифицировала 73.03% всех примеров.

AUC-ROC (Площадь под ROC-кривой):



AUC-ROC (Площадь под ROC-кривой): ROC (Receiver Operating Characteristic) - это график, который отображает производительность модели при различных порогах классификации. Площадь под ROC-кривой (AUC-ROC) измеряет общую производительность модели. Значение AUC-ROC близкое к 1 (ваше значение 0.8398) указывает на хорошую способность модели разделять классы.

4.Результаты дерево решений

Confusion Matrix:

[[35 14]

[15 25]]

Classification Report:

	precision	recall	f1-score	support
Class 0	0.70	0.71	0.71	49
Class 1	0.64	0.62	0.63	40

accuracy		0.67	89
macro avg	0.67	0.67	0.67 89
weighted avg	0.67	0.67	0.67 89

Accuracy: 0.6742
 Precision: 0.6410
 Recall: 0.6250
 F1-Score: 0.6329
 AUC-ROC: 0.6696

Матрица путаницы отображает количество верно и ошибочно классифицированных примеров для каждого класса.

Classification Report (Отчет о классификации):

Classification Report предоставляет детальные метрики для каждого класса (Class 0 и Class 1), а также средние значения (macro avg и weighted avg). В вашем случае:

Precision (точность) измеряет, как много из объектов, которые модель предсказала как положительные, действительно являются положительными. Precision для Class 0 составляет 0.70, а для Class 1 - 0.64.

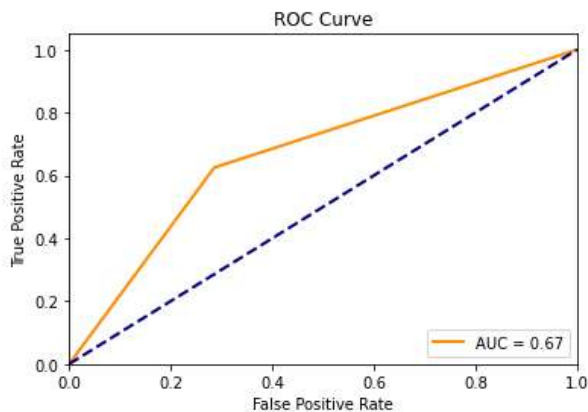
Recall (полнота) измеряет, как много из всех действительных положительных объектов модель правильно классифицировала. Recall для Class 0 составляет 0.71, а для Class 1 - 0.62.

F1-Score — это гармоническое среднее между точностью и полнотой. F1-Score для Class 0 составляет 0.71, а для Class 1 - 0.63.

Ассурасу (точность) - доля, верно, классифицированных примеров относительно общего числа примеров. Ассурасу составляет 0.6742, что означает, что модель правильно классифицировала 67.42% всех примеров.

AUC-ROC (Площадь под ROC-кривой):

AUC-ROC измеряет площадь под кривой ROC (Receiver Operating Characteristic), которая отображает производительность модели при различных порогах классификации. Значение AUC-ROC близкое к 1 (ваше значение 0.6696) указывает на способность модели разделять классы, но оно не очень близкое к 1, что может указывать на относительно низкую способность модели.



Эти метрики позволяют оценить производительность модели классификации, в частности, способность модели разделять классы,

сбалансированность точности и полноты, а также общую точность модели. Модель показала средние результаты в оценке классификации.

4. Дискуссия.

Машинное обучение может увеличить точность диагностики диабета и классификации его типа. Позволяет создавать индивидуальные планы лечения, учитывая уникальные характеристики каждого пациента. Модели машинного обучения могут предсказывать вероятность осложнений, что позволяет принимать меры по их предотвращению. Способствует интеграции данных из различных источников, обогащая информацию о пациенте. Системы поддержки принятия решений на основе машинного обучения помогают врачам и пациентам принимать более обоснованные решения.

Сбор и анализ медицинских данных поднимает вопросы конфиденциальности и безопасности данных пациентов. Машинное обучение требует доступа к большим и качественным данным, что может быть вызовом в медицинской практике. Многие модели машинного обучения, особенно нейронные сети, могут быть сложными для интерпретации, что усложняет объяснение решений пациентам и врачам. Машинное обучение требует специализированной экспертизы и обучения медицинского персонала для правильного применения.

Важно разрабатывать правила и нормативы, регулирующие применение машинного обучения в медицине, чтобы обеспечить прозрачность и ответственность. Необходимо обеспечивать защиту данных пациентов и соблюдать нормы конфиденциальности. Важным аспектом дискуссии является то, как машины могут работать с врачами и облегчать им принятие решений, а не заменять их. Подготовка медицинского персонала и специалистов в области машинного обучения становится все более важной, чтобы обеспечить правильное применение технологий.

5. Заключение.

Важным является правильное сбалансированное использование машинного обучения с учетом этических и практических аспектов. Машинное обучение позволяет диагностировать диабет с высокой точностью и рано выявлять риски развития заболевания. Это способствует раннему началу лечения и предотвращению осложнений.

Модели машинного обучения могут предсказывать вероятность осложнений и предупреждать врачей и пациентов. Это способствует более активному мониторингу и уходу. Машинное обучение представляет собой мощный инструмент для борьбы с диабетом, и его применение продолжит развиваться, создавая новые возможности для улучшения ухода за пациентами и научных исследований в этой области. Необходимо совмещать технологический прогресс с высокими стандартами этики и безопасности, чтобы обеспечить максимальную выгоду для пациентов и общества.

Литература

- [1] Pethunachiyar GA. Classification of diabetes patients using kernel based support vector machines. In: 2020 International Conference on Computer Communication Informatics (ICCCI). Coimbatore: IEEE (2020). p. 1–4. doi: 10.1109/ICCCI48352.2020.9104185
- [2] Gupta S, Verma HK, Bhardwaj D. Classification of diabetes using naïve bayes and support vector machine as a technique. In: Sachdeva A, Kumar P, Yadav OP, Garg RK, Gupta A, editors. Operations Management and Systems Engineering. Singapore: Springer (2021). p. 365–76. doi: 10.1007/978-981-15-6017-0_24
- [3] Рашка, С. Python и машинное обучение [Текст] / С. Рашка. – М. : ДМК Пресс, 2017. – 418 с.
- [4] Khattak A, Habib A, Asghar MZ, Subhan F, Razzak I, Habib A. Applying deep neural networks for user intention identification. *Soft Comput.* (2021) 25:2191–220. doi: 10.1007/s00500-020-05290-z
- [5] Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). *Foundations of Machine Learning*.// USA, Massachusetts: MIT Press
- [6] Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.* (2018) 19:1236–46. doi: 10.1093/bib/bbx044
- [7] Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Текст] / П. Флах.– М: ДМК Пресс, 2015. – 400 с.
- [8] Chen, L, Magliano, DJ and Zimmet, PZ. (2011) The worldwide epidemiology of type 2 diabetes mellitus-present and future perspectives.// *Nat Rev Endocrinol* 8: 228-236
- [9] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care—Addressing Ethical Challenges // *New England Journal of Medicine*, 378(11), 981-983.
- [10] Butt UM, Letchmunan S, Ali M, Hassan FH, Baqir A, Sherazi HHR. Machine learning based diabetes classification and prediction for healthcare applications. *J Healthcare Eng.* (2021) 2021:9930985. doi: 10.1155/2021/9930985
- [11] Alpaydin, Ethem (2010). *Introduction to Machine Learning*.// London: The MIT Press.
- [12] The Checkup. Diabetes Statistics: Read the Facts. (2020). Available online at: <https://www.singlecare.com/blog/news/diabetes-statistics/> (accessed December 20, 2021).
- [13] Qawqzeh YK, Bajahzar AS, Jemmali M, Ootom MM, Thaljaoui A. Classification of diabetes using photoplethysmogram (PPG) waveform analysis: logistic regression modeling. *Biomed Res Int.* (2020) 2020:3764653. doi: 10.1155/2020/3764653
- [14] Choubey DK, Kumar M, Shukla V, Tripathi S, Dhandhanika VK. Comparative analysis of classification methods with PCA and LDA for diabetes. *Curr Diabetes Rev.* (2020) 16:833–50. doi: 10.2174/1573399816666200123124008
- [15] Ahmad H, Asghar MU, Asghar MZ, Khan A, Mosavi AH. A hybrid deep learning technique for personality trait classification from text. *IEEE Access.* (2021) 9:146214–32. doi: 10.1109/ACCESS.2021.3121791
- [16] Alghazzawi D, Bamasaq O, Ullah H, Asghar MZ. Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection. *Appl Sci.* (2021) 11:11634. doi: 10.3390/app112411634
- [17] Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. *Proc Comput Sci.* (2019) 165:292–9. doi: 10.1016/j.procs.2020.01.047.

Мухамедиева Дилноз Тулкуновна

Д.т.н., профессор кафедры цифровые технологии и искусственный интеллект (ЦТИИ) Ташкентский институт инженеров ирригации и механизации сельского хозяйства (ТИИИМСХ)
E-mail: dilnoz134@rambler.ru

Раупова Мохинур Хайдар кизи

Преподаватель кафедры алгебра и математических анализ (АМА) Чирчикский государственный педагогический университет (ЧДПУ)
E-mail: r.mokhinur@gmail.com

Muhammedieva D.T., Raupova M.H.
Solving diabetes diagnosis problems using machine learning

Diagnosis of diabetes is an important task in medical practice, given the global increase in the number of patients with this chronic disease. Machine learning, with its ability to analyze and extract patterns from large volumes of data, provides a powerful tool for improving diabetes diagnosis and management. In this context, machine learning is being applied in various aspects of diabetes diagnosis, from early diagnosis to personalized treatment and monitoring. This opens up new opportunities for more effective prevention and management of this disease. In this paper, we will look at the various applications of machine learning in tasks related to the diagnosis of diabetes, and also highlight the role of machine learning in enhancing modern medical practice and improving the quality of life of patients with diabetes.

Keywords: diabetes, machine learning, model, Diabetes dataset, classification.