

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

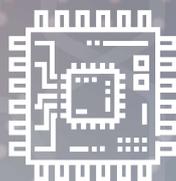
DESCENDANTS OF MUHAMMAD  
AL-KHWARIZMI  
SCIENTIFIC-PRACTICAL AND  
INFORMATION-ANALYTICAL JOURNAL



3(29)/2024

ISSN-2181-9211

MUHAMMAD AL-XORAZMIY NOMIDAGI  
TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI



# MUHAMMAD AL-XORAZMIY AVLODLARI

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

3(29)/2024

## 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. – Tahririyat kengashi raisi o'rinbosari

Raximov B.N. – t.f.d., prof. bosh muharrir

Nosirov X.X. – DSc., dots. 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.

Muradova A.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

Xudaybergenov J.D. – texnik muxarrir

Kengesbayev S.K. – texnik muxarrir

## MUNDARIJA

### DASTURIY VA KOMPYUTER INJINIRING TEKNOLOGIYALARINING ZAMONAVIY MUAMMOLARI

<b>Zayniddinov X.H., Xodjaeva D.F., Xuramov L.Y.</b> Разработка моделей и алгоритмов оптимального управления системами отопления и горячего водоснабжения.....	3
<b>Saidov A.A., Sharipov Sh.O.</b> Korporativ logistik axborot tizimlari ma'lumotlarini qayta ishlashda predmet-sohaga yo'naltirilgan ontologiyani qurish amaliyoti.....	8
<b>Babadjanov E.S.</b> Aqli chorva fermalarini sakllantirishda qoramol kasalliklarini aniqlash zamonaviy uslubiyati va algoritmik ta'minoti.....	16
<b>Kerimov K.F., Azizova Z.I.</b> Metodika analiza riskov informatsionnoy bezopanosti informatsionnykh resursov pri fizicheskikh угрозах.....	28
<b>Axatov A.R., Eshtemirov B.Sh.</b> Video ma'lumotlar asosida yo'llarda tirbandlik holatlarini aniqlashning intellektual algoritmlari.....	35
<b>Gulyamov Sh.M., Doshchanova M.Yu., Ruzibayev O.B.</b> Optimization of the control of the primary oil refining unit in a fuzzy environment.....	40
<b>Elov J.B., Iskandarova Sh.O., Mirkarimov A.M.</b> Texnologiya usulidagi favqulodda vaziyatlarda ogohlantirish tizimlaridan xabar uzatilishini modellashtirish.....	47
<b>Bekkamov F.A.</b> Tavsiya etish tizimlari uchun foydalanuvchi profilini yaratish, modellashtirish va shaxsiylashtirish.....	51
<b>Botirov F.B., Haydarov E.D., Gafurov A.A.</b> Milliy segmentga bo'ladigar DDOS hujumlarni aniqlash usuli va algoritmi.....	58
<b>Ruzibaev O.B., Doshanova M.Yu., Murodov D.D.</b> Эффективность использования различных библиотек Python для мультиклассовой классификации.....	62
<b>Muxamediyeva D.T., Raupova M.X.</b> Применение гибридных методов машинного обучения для диагностики диабета.....	71
<b>Geldibayev B.Y., Turmuxanov N.K., Yelmuratov Q.Q.</b> Chorva mollarida oqsolikli erta aniqlashda tayanch vektor mashinalari algoritmini qo'llash	79
<b>Baxriddinov A.Q.</b> Parabolik tipli tenglamaga qo'yilgan chegaraviy masalani differensial haydash usulida yechishning parallel hisoblash algoritmi.....	84

### OPTIK ALOQA TIZIMLARI, TELEKOMMUNIKATSIYA TARMOQLARI VA KOMMUTATSIYA TIZIMLARINING RIVOJLANISH TAMOYILLARI

<b>Лазарев А., Рейпназаров Е., Жолымбетова Э.</b> Анализ средств имитационного моделирования сетей VANET.....	88
<b>Usmanova N.B., G'ayratov Z.K.</b> Raqamli iqtisodiyot sharoitida aqli muhitni shakllantirish masalalari: infratuzilma va texnologik imkoniyatlar kesimida tahlil.....	104
<b>Matqurbonov D.M.</b> Модель системы управления маршрутизацией с использованием нечетких множеств.....	114
<b>F.K. Tojiyeva, U.R. Khamdamov, R.P. Abdurakhmanov.</b> Machine Learning Algorithms Analysis For Network traffic classification.....	119
<b>Allamuratova Z.J., Muxammedinov K.K.</b> OpenFlow protokolidan foydalangan holda dasturiy ta'minot bilan aniqlangan tarmoq bilan tarmoq boshqaruvini takomillashtirish.....	123
<b>I. Siddikov, H. Khujamatov, A. Temirov.</b> EDGE-FOG-CLOUD computing technologies in data processing in IoT-based energy supply system.....	127
<b>Lazarev A.P.</b> LoRa texnologiyasining axborot uzatish jarayoni modeli...	133
<b>Qodirov A.A.</b> Tarmoqning fizikaviy pog'onasi jarayonlarini muvofiqlashtirishning analitik modeli.....	141

### RAQAMLI TELEVIDENIYE VA RADIOESHITTIRISH, SIMSIZ TEKNOLOGIYALAR VA RADIOTEKNIKANI RIVOJLANTIRISH ISTIQBOLLARI

<b>Kuchkorov T.A., Sabitova N.Q.</b> Endoskopik tasvirlarning sifatini baholash usullari.....	145
<b>Писецкий Ю.В., Вотинков К.А.</b> Перспективы использования множественного доступа в системах радиосвязи.....	148

УДК 519.71

Мухамедиева Д.Т., Раупова М.Х.

## Применение гибридных методов машинного обучения для диагностики диабета

Гибридные методы машинного обучения представляют собой мощный инструмент для улучшения точности диагностики сахарного диабета, одного из наиболее распространенных хронических заболеваний. В работе исследуется применение ансамблевых методов, таких как случайный лес, градиентный бустинг и бэггинг для диагностики сахарного диабета и анализирует их преимущества и вызовы. Гибридные методы способствуют увеличению точности диагностики и снижению ложных срабатываний и ложноотрицательных результатов. Они позволяют работать с разнородными данными, обеспечивают устойчивость к переобучению и предоставляют информацию о важности признаков. В целом, гибридные методы машинного обучения представляют собой перспективный инструмент для улучшения диагностики сахарного диабета и могут способствовать более эффективному выявлению и управлению этим хроническим заболеванием. Дальнейшие исследования и разработки в этой области могут привести к созданию более точных и надежных методов диагностики и лечения сахарного диабета.

**Ключевые слова:** гибрид, машинное обучение, диабет, диагностика, модель.

### 1. Введение.

Актуальность применения машинного обучения в области диагностики и управления диабетом трудно переоценить. Машинное обучение позволяет создавать модели, которые могут анализировать множество медицинских данных пациентов и предсказывать риск развития диабета. Это позволяет врачам и пациентам принимать меры по ранней диагностике и профилактике заболевания, позволяет анализировать характеристики и реакцию каждого пациента на лечение. Это помогает разрабатывать индивидуальные планы лечения и управления диабетом, что повышает эффективность терапии. Системы мониторинга глюкозы и других биометрических параметров становятся все более доступными и помогают анализировать эти данные в реальном времени и предупреждать о вероятных проблемах или необходимости коррекции лечения. Машинное обучение способствует интеграции данных из различных источников, таких как медицинские карты, данные лабораторных исследований, информация о лекарствах и мониторинге, что позволяет врачам и исследователям получить более полное представление о состоянии пациентов. Автоматизированные системы и инструменты машинного обучения помогают врачам принимать более обоснованные решения и предоставлять более точный и персонализированный уход для пациентов с диабетом. Все эти аспекты подчеркивают важность машинного обучения в области диабета и делают его актуальным и перспективным инструментом для медицинской практики и исследований [1].

Машинное обучение позволяет разрабатывать модели классификации, которые могут предсказывать, имеет ли пациент диабет или нет, и, при необходимости, классифицировать тип диабета (тип 1 или тип 2). Для этого используются наборы данных с клиническими и лабораторными показателями, такими как уровень глюкозы в крови, индекс массы тела (ИМТ), артериальное давление и другие факторы риска. Модели машинного обучения

могут помочь в ранней диагностике диабета и оценке риска развития диабета у лиц с предрасположенностью к этому заболеванию. Они могут анализировать большие объемы данных пациентов и предупреждать врачей о возможности диабета, что позволяет начать лечение на ранних стадиях. Существует несколько наборов данных для машинного обучения, которые используются для задач диагностики диабета и прогнозирования его развития. Вот некоторые из наиболее известных наборов данных [2]:

**Pima Indians Diabetes Database:** Этот набор данных содержит информацию о пациентах из племени папьян-индейцев, включая такие признаки, как уровень глюкозы в крови, артериальное давление, ИМТ и другие. Он часто используется для задач классификации диабета.

**Diabetes Dataset (UCI):** Набор данных UCI Diabetes содержит информацию о пациентах, включая 8 признаков, такие как возраст, ИМТ, уровень глюкозы и другие. Этот набор данных также широко используется в машинном обучении для задач классификации диабета.

**Diabetes dataset (Scikit-Learn):** Это набор данных о диабете, включенный в библиотеку scikit-learn. Он содержит информацию о пациентах и их признаках, таких как уровень глюкозы, артериальное давление и другие.

**Indian Diabetes Dataset:** Этот набор данных содержит информацию о пациентах из Индии и включает признаки, такие как возраст, пол, количество беременностей и другие.

**Diabetes Retinopathy Debrecen Data Set:** Этот набор данных сосредотачивается на диагностике ретинопатии диабетической, оценке степени ее развития. Он включает в себя изображения сетчатки глаз пациентов.

Эти наборы данных предоставляют различные типы информации о пациентах и их состоянии, и их можно использовать для обучения моделей машинного обучения с целью диагностики диабета и прогнозирования его развития.

Датасет Diabetes представляет собой один из наиболее распространенных и широко используемых наборов данных для задач машинного обучения и статистики. Этот набор данных предоставляет информацию о пациентах, страдающих диабетом, и характеристиках их здоровья. Содержит информацию о 442 пациентах. Каждый пациент описывается одной записью в наборе данных. В этом наборе данных содержатся 10 числовых признаков, которые характеризуют состояние пациентов. Важными признаками являются:

Уровень глюкозы в крови (blood sugar level).

Артериальное давление (blood pressure).

Индекс массы тела (body mass index, BMI).

Индекс уровня молекулы сердечного заболевания (serum levels of a heart disease marker).

Эти признаки имеют важное значение для диагностики и управления диабетом.

Цель применения ансамблевых методов машинного обучения для диагностики диабета заключается в улучшении процесса диагностики сахарного диабета путем повышения точности, надежности и эффективности выявления этого хронического заболевания [3].

Задачи машинного обучения в контексте диабета могут быть разнообразными и охватывать различные аспекты диагностики, управления и исследования этого заболевания. Вот некоторые из ключевых задач, которые машинное обучение может решать в области диабета [4-7]:

Классификация диабета: Одной из основных задач является классификация типа диабета (тип 1 или тип 2) или даже различных стадий диабета. Модели машинного обучения могут анализировать клинические и лабораторные данные пациентов для правильной классификации.

Ранняя диагностика: Создание моделей для ранней диагностики диабета и предсказания риска его развития. Это позволяет начать лечение на ранних стадиях, когда оно наиболее эффективно.

Прогнозирование прогрессии: Задачей может быть прогнозирование темпов прогрессии диабета и определение, какие пациенты подвержены более быстрому ухудшению состояния.

Оптимизация лечения: Машинное обучение может помочь врачам оптимизировать планы лечения для каждого пациента, учитывая его характеристики и реакцию на лекарства.

Мониторинг и предупреждение осложнений: Задачей может быть создание систем мониторинга для предупреждения осложнений, таких как гипогликемия, диабетическая кетоацидоз и диабетическая ретинопатия.

Интеграция медицинских данных: Машинное обучение может помочь в интеграции данных из различных источников, чтобы врачи и исследователи могли получить более полное представление о состоянии пациентов.

Исследование новых методов лечения: Машинное обучение может использоваться для

анализа данных клинических исследований и 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: Машинное обучение может помочь оценить, насколько эффективно лечение действует на конкретного пациента и, при необходимости, скорректировать план лечения.

Улучшение качества ухода: Основной задачей является повышение качества ухода для пациентов с диабетом, снижение риска осложнений и улучшение их качества жизни.

Эти 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.

Применение ансамблевых методов машинного обучения для диагностики сахарного диабета может улучшить точность и надежность моделей. Ансамблевые методы объединяют несколько базовых моделей в одну, чтобы уменьшить переобучение, повысить обобщающую способность и улучшить качество прогнозов [8].

1. Случайный лес (Random Forest) - это мощный ансамблевый метод машинного обучения, который используется для задач классификации и регрессии. Он основан на идее создания нескольких деревьев решений и объединения их результатов для улучшения точности и устойчивости модели [9].

Случайный лес состоит из множества деревьев решений, которые обучаются на различных подмножествах данных. Каждое дерево может быть слабой моделью, но их комбинация создает сильную модель. При построении каждого дерева, случайный лес случайным образом выбирает подмножество признаков для разделения данных. Это помогает уменьшить корреляцию между деревьями и увеличить разнообразие моделей. Для обучения каждого дерева используется бутстрэп выборка (выборка с повторением) из исходных данных. Это позволяет учесть разнообразие данных. После обучения всех деревьев, случайный лес агрегирует их прогнозы. В задаче классификации, чаще всего используется голосование большинства, а в задаче регрессии - усреднение прогнозов.

Случайный лес часто демонстрирует высокую точность классификации и регрессии. Благодаря случайному выбору признаков и использованию бутстрэп выборки, случайный лес устойчив к переобучению. Можно использовать случайный лес для обработки данных с большим количеством признаков. Случайный лес позволяет оценить

важность каждого признака для задачи. Метод применим к разнообразным типам данных.

2. Градиентный бустинг - это ансамбль, в котором каждое дерево настраивается на ошибки предыдущего дерева. Это улучшает качество предсказаний и позволяет работать с данными разного типа. Примеры включают градиентный бустинг с деревьями (Gradient Boosting Trees) и библиотеки, такие как XGBoost, LightGBM и CatBoost. Градиентный бустинг (Gradient Boosting) - это мощный ансамблевый метод машинного обучения, который также используется для задач классификации и регрессии. В отличие от бэггинга, где базовые модели обучаются независимо, градиентный бустинг строит ансамбль моделей последовательно, при этом каждая новая модель настраивается на ошибки предыдущей модели. Градиентный бустинг строит ансамбль моделей в последовательности. Новая модель обучается так, чтобы исправить ошибки предыдущей модели. Использует метод градиентного спуска для настройки параметров моделей. Это позволяет оптимизировать функцию потерь. В качестве базовых моделей в градиентном бустинге часто используются слабые модели, такие как деревья решений небольшой глубины. Это делает метод более устойчивым к переобучению. Градиентный бустинг уделяет особое внимание ошибкам предыдущей модели, участвующей в обучении следующей модели. Это позволяет создавать модели, которые сосредотачиваются на сложных примерах [10].

Градиентный бустинг обычно обеспечивает высокую точность предсказаний. Последовательное обучение и внимание к ошибкам предыдущих моделей делают градиентный бустинг устойчивым к переобучению. Градиентный бустинг может использоваться с различными типами данных, включая числовые и категориальные признаки. Позволяет оценить важность каждого признака для задачи. Градиентный бустинг нашел применение в различных областях машинного обучения, включая соревнования по анализу данных.

3. Бэггинг — это метод, при котором несколько моделей обучаются независимо на разных подмножествах данных, а затем их прогнозы усредняются. Бэггинг (Bootstrap Aggregating) — это метод ансамблевого машинного обучения, в котором несколько моделей (например, деревьев решений или любых других базовых моделей) обучаются независимо на различных подмножествах данных, а затем их прогнозы усредняются или объединяются. Этот метод создает ансамбль моделей, который часто более стабилен и точен, чем отдельные модели. При использовании бэггинга каждая модель обучается на случайной выборке (с повторениями) из исходных данных. Это означает, что некоторые примеры могут вообще не попадать. После обучения всех моделей их прогнозы усредняются для получения финального прогноза. В задаче классификации, чаще всего

используется голосование большинства, а в задаче регрессии - усреднение прогнозов. Бэггинг уменьшает переобучение, потому что каждая модель обучается на разных подмножествах данных, что способствует устойчивости модели [11].

Благодаря усреднению прогнозов нескольких моделей, бэггинг снижает дисперсию ошибки. Бэггинг способствует созданию более устойчивых и обобщающих моделей. Применимость к различным базовым моделям: Бэггинг может использоваться с различными типами базовых моделей. Из-за случайного выбора данных бэггинг менее чувствителен к выбросам в данных. Модели могут обучаться независимо, что позволяет использовать параллельные вычисления.

Примеры бэггинга включают в себя метод случайного леса (Random Forest) и Bagged Decision Trees. Эти методы нашли широкое применение в различных областях машинного обучения и анализа данных для улучшения качества и надежности моделей.

4. Адаптивный бустинг (Adaptive Boosting или AdaBoost) — это алгоритм машинного обучения, который используется для улучшения производительности моделей классификации. Он работает путем комбинирования нескольких слабых классификаторов в одну сильную модель. AdaBoost использует слабые классификаторы, которые могут быть моделями, недостаточно мощными для решения задачи классификации в отдельности. Каждому образцу в обучающем наборе присваивается вес, и в начале обучения они равны. Алгоритм сосредотачивается на образцах, которые неправильно классифицированы предыдущими моделями. AdaBoost адаптирует веса образцов на каждом этапе обучения. Ошибки классификации увеличивают веса неправильно классифицированных образцов, что позволяет следующим моделям сосредотачиваться на этих образцах. Комбинирует предсказания всех слабых моделей, взвешивая их на основе их производительности. Это создает сильную модель, которая может быть более точной, чем каждая из слабых моделей в отдельности. Алгоритм AdaBoost выполняется в несколько итераций (этапов), и на каждом этапе добавляется новая слабая модель, которая фокусируется на ошибках, допущенных предыдущими моделями [12].

5. Смешивание моделей (Model Stacking).

Модель стекинга — это метод, при котором несколько моделей объединяются, и их прогнозы становятся входными данными для финальной модели. Это может быть полезным для объединения различных типов моделей.

Смешивание моделей, также известное как "Model Stacking" или "Ensemble Stacking", представляет собой метод ансамблирования в машинном обучении, в котором результаты нескольких различных моделей комбинируются с целью улучшения общей производительности. Этот

метод широко используется для решения сложных задач классификации и регрессии [13-14].

В начале создаются несколько различных базовых моделей (или классификаторов), использующих разные алгоритмы или параметры. Каждая базовая модель обучается на одном и том же обучающем наборе данных. Они могут использовать разные признаки или параметры, что увеличивает разнообразие моделей. Создается дополнительная модель (называемая "моделью стекинга" или "метамоделью"), которая принимает предсказания базовых моделей в качестве входных признаков. Метамодель обучается на выходах базовых моделей, предсказывая итоговый результат. Это позволяет метамодели объединить предсказания базовых моделей и учесть их сильные и слабые стороны.

После обучения метамодели она может быть использована для классификации или регрессии на новых данных. Преимущества модели стекинга включают повышение обобщающей способности, улучшение производительности и увеличение устойчивости предсказаний. Однако стекинг требует более сложной настройки и управления, и может потребовать больше вычислительных ресурсов.

Примеры моделей стекинга включают в себя использование разных алгоритмов машинного обучения, таких как решающие деревья, метод ближайших соседей, логистическая регрессия и даже нейронные сети, в качестве базовых моделей, а затем применение модели стекинга, такой как градиентный бустинг или случайный лес, для объединения их результатов.

### 3. Результат.

Чтобы применить ансамблевые методы к диагностике сахарного диабета, нужно следовать следующим шагам:

**Собрать и подготовить данные:** получите надежные и чистые данные о пациентах, включая признаки, такие как возраст, пол, ИМТ (индекс массы тела), уровень глюкозы в крови, артериальное давление и другие. Разделите данные на обучающий и тестовый наборы.

**Выбрать ансамблевый метод:** решите, какой из ансамблевых методов наилучшим образом подходит для вашей задачи классификации диабета.

**Обучить ансамбль:** обучите выбранный ансамбль на обучающем наборе данных. Тонкая настройка параметров может улучшить производительность.

**Оценить производительность:** Используйте метрики, такие как accuracy, precision, recall, F1-score и AUC-ROC, для оценки производительности модели на тестовом наборе данных.

**Валидация и настройка:** Проведите кросс-валидацию и тюнинг параметров, чтобы убедиться, что модель работает хорошо и не переобучается.

**Применение модели:** Используйте обученную модель для диагностики сахарного диабета на новых пациентах.

1. По разработанной программе по алгоритму Случайный лес (Random Forest) получены следующие результаты:

Confusion Matrix:

[[37 12]

[13 27]]

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: 0.8125

Метрики, представленные в отчете о классификации и матрице путаницы, оценивают производительность модели машинного обучения на задаче классификации, где есть два класса: "Class 0" и "Class 1". Давайте разберемся, что означает каждая из этих метрик:

**Confusion Matrix (Матрица путаницы):**

Первый ряд матрицы путаницы представляет истинные значения класса 0 (Class 0), а второй ряд - истинные значения класса 1 (Class 1).

Первый столбец матрицы показывает, сколько объектов было предсказано как класс 0, и сколько из них действительно принадлежат к классу 0.

Второй столбец показывает, сколько объектов было предсказано как класс 0, но они фактически принадлежат к классу 1.

Третий столбец показывает, сколько объектов было предсказано как класс 1, и сколько из них действительно принадлежат к классу 0.

Четвертый столбец показывает, сколько объектов было предсказано как класс 1, и сколько из них действительно принадлежат к классу 1.

Accuracy оценивает долю правильно классифицированных объектов по отношению ко всем объектам. В данном случае, точность равна 0.7191, что означает, что модель правильно классифицировала 71.91% объектов.

**Precision (Точность):**

Precision (точность) измеряет долю истинных положительных предсказаний относительно всех положительных предсказаний. В данном случае, точность для "Class 0" равна 0.74, а для "Class 1" - 0.69.

**Recall (Полнота):**

Recall (полнота) измеряет долю истинных положительных предсказаний относительно всех истинных положительных объектов. В данном

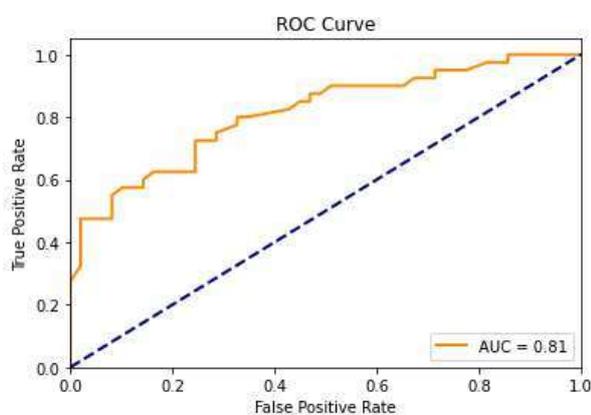
случае, полнота для "Class 0" равна 0.76, а для "Class 1" - 0.68.

F1-Score:

F1-мера является средним гармоническим между точностью и полнотой. Это числовая оценка баланса между точностью и полнотой. В данном случае, F1-мера для "Class 0" составляет 0.75, а для "Class 1" - 0.68.

AUC-ROC (Площадь под кривой ROC):

AUC-ROC - это показатель качества модели, который оценивает способность модели разделять классы и зависит от порогового значения вероятности. Значение 0.8125 означает, что модель хорошо разделяет классы, и AUC-ROC близка к 1, что свидетельствует о хорошей производительности модели.



В целом, метрики оценивают производительность модели в задаче классификации. Точность, полнота и F1-мера оценивают качество классификации для каждого класса, а AUC-ROC предоставляет общую оценку качества разделения классов.

2. По разработанной программе по алгоритму Градиентный бустинг (Gradient Boosting) получены следующие результаты:

Confusion Matrix:

```
[[37 13]
 [14 25]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.74	0.73	50
1	0.66	0.64	0.65	39
accuracy			0.70	89
macro avg	0.69	0.69	0.69	89
weighted avg	0.70	0.70	0.70	89

Accuracy: 0.6966292134831461  
 Precision: 0.6578947368421053  
 Recall: 0.6410256410256411  
 F1-Score: 0.6493506493506495  
 AUC-ROC: 0.7605128205128205

Метрики, представленные в отчете о классификации и матрице путаницы, оценивают

производительность модели машинного обучения в задаче бинарной классификации. В данном случае, у вас есть два класса: "0" и "1". Давайте разберемся, что означают эти метрики:

Confusion Matrix (Матрица путаницы):

Первый ряд матрицы путаницы представляет истинные значения класса "0", а второй ряд - истинные значения класса "1".

Первый столбец матрицы показывает, сколько объектов было предсказано как класс "0", и сколько из них действительно принадлежат к классу "0".

Второй столбец показывает, сколько объектов было предсказано как класс "0", но они фактически принадлежат к классу "1".

Третий столбец показывает, сколько объектов было предсказано как класс "1", и сколько из них действительно принадлежат к классу "0".

Четвертый столбец показывает, сколько объектов было предсказано как класс "1", и сколько из них действительно принадлежат к классу "1".

Accuracy (Точность):

Accuracy оценивает долю правильно классифицированных объектов по отношению ко всем объектам. В данном случае, точность равна 0.6966, что означает, что модель правильно классифицировала 69.66% объектов.

Precision (Точность):

Precision (точность) измеряет долю истинных положительных предсказаний относительно всех положительных предсказаний. В данном случае, точность для класса "0" равна 0.73, а для класса "1" - 0.66.

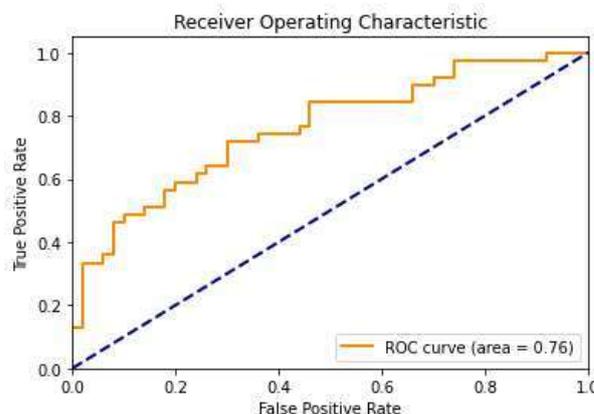
Recall (Полнота):

Recall (полнота) измеряет долю истинных положительных предсказаний относительно всех истинных положительных объектов. В данном случае, полнота для класса "0" равна 0.74, а для класса "1" - 0.64.

F1-Score:

F1-мера является средним гармоническим между точностью и полнотой. Это числовая оценка баланса между точностью и полнотой. В данном случае, F1-мера для класса "0" составляет 0.73, а для класса "1" - 0.65.

AUC-ROC (Площадь под кривой ROC):



AUC-ROC - это показатель качества модели, который

оценивает способность модели разделять классы и зависит от порогового значения вероятности. Значение 0.7605 означает, что модель хорошо разделяет классы, и AUC-ROC близка к 1, что свидетельствует о хорошей производительности модели.

### 3. Результат Бэггинг (Bagging):

В этой программе мы используем метод бэггинга с базовым классификатором, который является деревом решений (DecisionTreeClassifier). Мы разделяем данные на обучающий и тестовый наборы, обучаем модель бэггинга и оцениваем ее производительность, а также вычисляем матрицу путаницы, отчет о классификации и пять основных метрик классификации

Confusion Matrix:

```
[[39 11]
```

```
[15 24]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.72	0.78	0.75	50
1	0.69	0.62	0.65	39
accuracy		0.71		89
macro avg	0.70	0.70	0.70	89
weighted avg	0.71	0.71	0.71	89

Accuracy: 0.7078651685393258

Precision: 0.6857142857142857

Recall: 0.6153846153846154

F1-Score: 0.6486486486486486

AUC-ROC: 0.7892307692307693

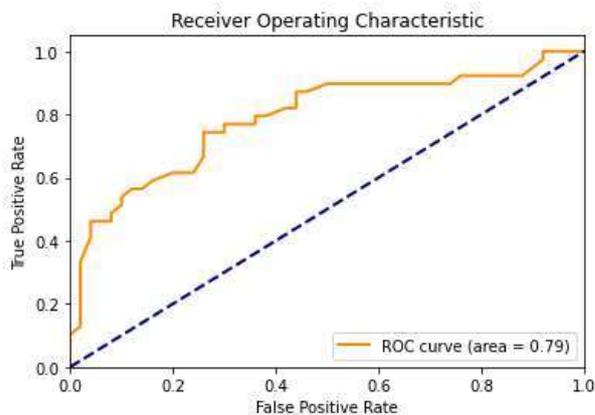
Accuracy оценивает долю правильно классифицированных объектов по отношению ко всем объектам. В данном случае, точность равна 0.7079, что означает, что модель правильно классифицировала 70.79% объектов.

Precision (точность) измеряет долю истинных положительных предсказаний относительно всех положительных предсказаний. В данном случае, точность для класса "0" равна 0.72, а для класса "1" - 0.69.

Recall (полнота) измеряет долю истинных положительных предсказаний относительно всех истинных положительных объектов. В данном случае, полнота для класса "0" равна 0.78, а для класса "1" - 0.62.

F1-мера является средним гармоническим между точностью и полнотой. Это числовая оценка баланса между точностью и полнотой. В данном случае, F1-мера для класса "0" составляет 0.75, а для класса "1" - 0.65.

AUC-ROC - это показатель качества модели, который оценивает способность модели разделять классы и зависит от порогового значения вероятности.



Значение 0.7892 означает, что модель хорошо разделяет классы, и AUC-ROC близка к 1, что свидетельствует о хорошей производительности модели классификации.

### 4. Результат Адаптивный бустинг (AdaBoost):

В разработанной программе мы используем метод AdaBoost с базовыми деревьями решений (DecisionTreeClassifier). Мы разделяем данные на обучающий и тестовый наборы, обучаем модель AdaBoost и оцениваем ее производительность, а также вычисляем матрицу путаницы, отчет о классификации и пять основных метрик классификации.

Confusion Matrix:

```
[[42 8]
```

```
[18 21]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.70	0.84	0.76	50
1	0.72	0.54	0.62	39
accuracy		0.71		89
macro avg	0.71	0.69	0.69	89
weighted avg	0.71	0.71	0.70	89

Accuracy: 0.7078651685393258

Precision: 0.7241379310344828

Recall: 0.5384615384615384

F1-Score: 0.6176470588235294

AUC-ROC: 0.698974358974359

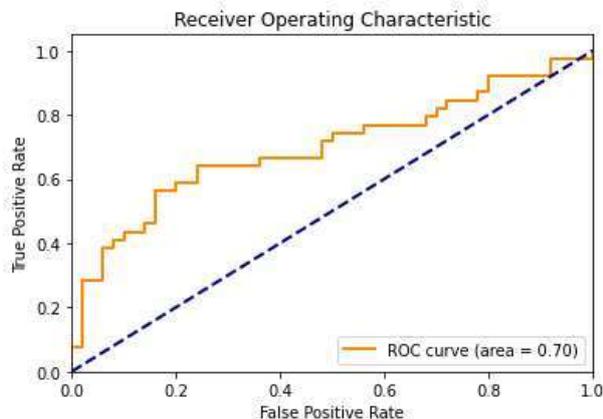
Accuracy оценивает долю правильно классифицированных объектов по отношению ко всем объектам. В данном случае, точность равна 0.7079, что означает, что модель правильно классифицировала 70.79% объектов.

Precision (точность) измеряет долю истинных положительных предсказаний относительно всех положительных предсказаний. В данном случае, точность для класса "0" равна 0.70, а для класса "1" - 0.72.

Recall (полнота) измеряет долю истинных положительных предсказаний относительно всех истинных положительных объектов. В данном случае, полнота для класса "0" равна 0.84, а для класса "1" - 0.54.

F1-мера является средним гармоническим между точностью и полнотой. Это числовая оценка баланса между точностью и полнотой. В данном случае, F1-мера для класса "0" составляет 0.76, а для класса "1" - 0.62.

AUC-ROC (Площадь под кривой ROC):



AUC-ROC - это показатель качества модели, который оценивает способность модели разделять классы и зависит от порогового значения вероятности. Значение 0.6989 означает, что модель разделяет классы, но с некоторой неопределенностью, и AUC-ROC близка к 0.5, что свидетельствует о средней производительности модели классификации.

5.Результат Смешивание моделей (Model Stacking):

Этот метод показывает, как можно создать стекинг моделей для задачи диагностики диабета, используя логистическую регрессию и случайный лес в качестве базовых моделей и градиентный бустинг в качестве метамодели. Вы можете изменить базовые модели и параметры в соответствии с вашими потребностями и данными.

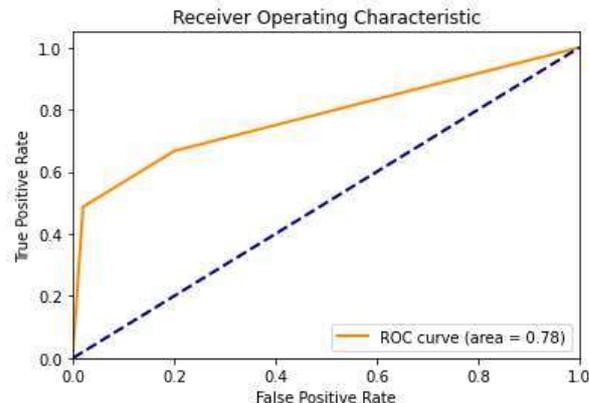
Ассигасу оценивает долю правильно классифицированных объектов по отношению ко всем объектам. В данном случае, точность равна 0.764, что означает, что модель правильно классифицировала 76.4% объектов.

Precision (точность) измеряет долю истинных положительных предсказаний относительно всех положительных предсказаний. В данном случае, точность для класса "0" равна 0.71, а для класса "1" - 0.95.

Recall (полнота) измеряет долю истинных положительных предсказаний относительно всех истинных положительных объектов. В данном случае, полнота для класса "0" равна 0.98, а для класса "1" - 0.49.

F1-мера является средним гармоническим между точностью и полнотой. Это числовая оценка баланса между точностью и полнотой. В данном случае, F1-мера для класса "0" составляет 0.82, а для класса "1" - 0.64.

AUC-ROC (Площадь под кривой ROC):



AUC-ROC - это показатель качества модели, который оценивает способность модели разделять классы и зависит от порогового значения вероятности. Значение 0.7756 означает, что модель разделяет классы с неплохой производительностью, и AUC-ROC близка к 0.5, что свидетельствует о средней производительности модели классификации.

Наилучший результат показал случайный лес.

#### 4. Дискуссия.

Применение ансамблевых методов машинного обучения для диагностики сахарного диабета представляет собой важную область исследования и практического применения в медицине.

Однако следует отметить, что применение ансамблевых методов также сопряжено с некоторыми вызовами и вопросами, такими как необходимость большего объема данных для обучения сложных моделей и требования к вычислительным ресурсам. Также важно обеспечить безопасность и конфиденциальность медицинских данных при их использовании в алгоритмах машинного обучения.

В целом, ансамблевые методы машинного обучения играют важную роль в улучшении диагностики сахарного диабета и могут привести к более эффективным и точным методам выявления этого заболевания.

#### 5. Заключение.

Применение ансамблевых методов машинного обучения для диагностики сахарного диабета представляет собой многообещающую и важную область исследований и практического применения в медицине. В ходе нашей дискуссии мы рассмотрели ряд ключевых моментов, связанных с использованием ансамблевых методов для этой цели.

Важно подчеркнуть, что ансамблевые методы машинного обучения, такие как случайный лес, градиентный бустинг, бэггинг и адаптивный бустинг, предоставляют значительные преимущества при диагностике сахарного диабета. Ансамблевые методы объединяют предсказания нескольких моделей, что часто приводит к улучшению точности диагностики. Это особенно важно в медицинской сфере, где даже небольшие улучшения могут спасти жизни пациентов. Позволяют снизить количество ложноположительных и ложноотрицательных результатов диагностики, что является критически

важным для эффективного лечения пациентов. Ансамблевые методы могут объединять информацию из различных источников, таких как клинические данные, биохимические показатели и изображения, что делает диагностику более надежной. Обладают хорошей устойчивостью к переобучению, что особенно важно при работе с ограниченными объемами медицинских данных. Некоторые ансамблевые методы, такие как случайный лес, могут предоставить информацию о важности признаков, что помогает врачам и исследователям лучше понимать факторы, влияющие на диагностику диабета. В итоге, применение ансамблевых методов машинного обучения для диагностики сахарного диабета имеет огромный потенциал для улучшения здравоохранения и качества жизни пациентов. Дальнейшие исследования и разработки в этой области будут способствовать развитию эффективных инструментов для диагностики и лечения сахарного диабета.

#### Литература

- [1] 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
- [2] 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
- [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, Bajazhar AS, Jemmali M, Otoom 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

#### 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](mailto: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](mailto:r.mokhinur@gmail.com)

#### D.T. Muhammedieva, M.H. Raupova.

#### Application of hybrid machine learning methods for diabetes diagnosis

Hybrid machine learning techniques provide a powerful tool for improving the diagnostic accuracy of diabetes mellitus, one of the most common chronic diseases. The paper explores the use of ensemble methods such as random forest, gradient boosting and bagging for diagnosing diabetes mellitus and analyzes their advantages and challenges. Hybrid methods help to increase diagnostic accuracy and reduce false positives and false negatives. They allow you to work with heterogeneous data, provide resistance to overfitting, and provide information about the importance of features. Overall, hybrid machine learning methods represent a promising tool for improving diabetes diagnosis and may contribute to more effective detection and management of this chronic disease. Further research and development in this area may lead to more accurate and reliable methods for diagnosing and treating diabetes.

**Keywords:** hybrid, machine learning, diabetes, diagnostics, model.

Geldibayev B.Y., Turmuxanov N.K., Yelmuratov Q.Q.

## Chorva mollarida oqsoqlikni erta aniqlashda tayanch vektor mashinalari algoritmini qo'llash

Mazkur tadqiqot ishida chorva mollarida oqsoqlik kasalligini mashinali o'qitish algoritmlari yordamida erta aniqlash masalasi qarab chiqilgan. Bunda chorva mollarining oyog'iga o'rnatilgan pedometrlar yordamida ularning kunlik faolligi haqidagi ma'lumotlar yig'ib borilgan va to'plangan ma'lumotlarni dastlabki qayta ishlash orqali o'quv tanlanma hosil qilingan. Shakllantirilgan o'quv tanlanma tayanch vektor mashinalari algoritmi yordamida o'qitish amalga oshirilgan.

**Kalit so'zlar:** mashinali o'qitish, ma'lumotlar tahlili, belgilarni ajratib olish, tayanch vektor mashinalari, algoritm.

### Kirish

Oxirgi yillarda davlatimizda chorvachilik sohasida zamonaviy innovatsion usullarni joriy etish, mahsulot ishlab chiqarish hajmini va samaradorligini oshirish, chorvachilik subyektlarini davlat tomonidan qo'llab-quvvatlashga qaratilgan bir qator qarorlar qabul qilinmoqda [1, 2]. Natijada hududlarda chorva mahsulotlarini ishlab chiqarish bilan shug'ullanadigan fermalar soni yildan-yilga ortib bormoqda.

Fermalar sonining ortishi o'z navbatida ularning samaradorligini oshirish bo'yicha innovatsion yechimlarni talab etadi. Buning uchun hozirda jahon tajribasida asosan zamonaviy axborot texnologiyalari va mashinali o'qitish algoritmlarining imkoniyatlari keng qo'llanilmoqda [3, 4]. Hususan sohada sensorlar va videokuzatuv asosida ma'lumotlarni to'plash va ularga mashinali o'qitish algoritmlari yordamida ishlov berish orqali qarorlar qabul qilish tizimlarini joriy etish kabi innovatsion yechimlarni qo'llash soha samaradorligini sezilarli darajada oshirishga yordam beradi [5].

Hozirgi vaqtga kelib sut va sut mahsulotlari is'temoli yildan yilga ortib bormoqda. Hisob-kitoblarga ko'ra, sut va sut mahsulotlari iste'molchilari bazasi 2009 yildagi 1,8 milliard odamdan 2030 yilga kelib 4,9 milliardga ko'tariladi [6]. Shundan kelib chiqqan holda sut ishlab chiqarish samaradorligini oshirish dolzarb masalaga aylanib bormoqda. Shu o'rinda sut ishlab chiqarish sohasining samaradorligiga ta'sir ko'rsatadigan asosiy muammolardan biri bu sog'in sigirlarda uchraydigan kasalliklar bo'lib ular chorva mollarining farovonligining pasayishiga va sut ishlab chiqarish hajmining sezilarli darajada kamayishiga olib keladi [7, 8].

Chorva mahsulotlarini ishlab chiqarish samaradorligi saqlanayotgan chorva mollarining farovonligi va sog'ligi bilan to'g'ridan-to'g'ri bog'liq bo'lib hisoblanadi. Ularda uchraydigan turli hil kasalliklar mahsulodkorlikning sezilarli darajada pasayishiga olib keladi [9]. Asosan sut mahsulotlarini ishlab chiqarishga mo'ljallangan chorva mollarida mastit va oqsoqlik kabi kasalliklar eng keng tarqalgan kasalliklardan bo'lib iqtisodiy jihatdan fermerlarga jiddiy zararlar keltiradi [10, 11]. Bulardan oqsoqlik kasalligi o'zining tarqalganligi va uni davolash bo'yicha qilinadigan sarf xarajatlar miqdori bo'yicha mastitdan keyingi ikkinchi o'rinda turadi [12]. Kasallikni erta aniqlash va kerakli choralarini o'z vaqtida qo'llash orqali jarayon uchun qilinadigan sarf xarajatlarni kamaytirishga yordam beradi [13]. Oqsoqlik kasalligini erta aniqlash

masalasi bo'yicha olib borilgan tadqiqot ishlarida jarayonda pedometrlarni qo'llash taklif etilgan va tegishli dasturiy majmua arxitekturasini loyihalash amalga oshirilgan [14]. Quyida ushbu arxitektura asosida qurilgan tizim orqali ma'lumotlarni yig'ish, ularga dastlabki ishlov berish va mashinali o'qitish algoritmlari yordamida qayta ishlash masalasi qarab chiqilgan.

### Ma'lumotlarni to'plash

Ma'lumotlarni to'plash jarayoni ilmiy tadqiqotning asosiy tarkibiy qismidir. To'plangan ma'lumotlarning sifati va to'g'riligi tahlil natijalariga va keyingi xulosalarga bevosita ta'sir qiladi. Tobora aniq va batafsil ma'lumotlarni talab qiladigan tez o'zgaruvchan bugungi dunyoda ma'lumotlarni to'plash jarayonini to'g'ri tashkil etish va amalga oshirish muhim ahamiyatga ega.

Yuqorida keltirilgan tizimning asosiy maqsadi ma'lumotlarni to'plashdan iborat bo'lib u quyidagicha tarkibiy qismlardan iborat:

1. Ma'lumotlarni yig'ish dasturiy moduli;
2. Ma'lumotlarni dastlabki qayta ishlash dasturiy moduli;
3. Ma'lumotlarni markazlashtirilgan saqlash moduli;
4. Tahlil va qaror qabul qilish dasturiy moduli;
5. Foydalanuvchi interfeysi dasturiy moduli;
6. Xavfsizlik dasturiy moduli;
7. Boshqa tizimlar bilan integratsiya dasturiy moduli.

Tizimda ma'lumotlarni to'plash jarayoniga puxta e'tibor qaratilgan bo'lib bunda zamonaviy sensor texnologiyalarining imkoniyatlaridan keng foydalanilgan. Bunda asosiy qurilma pedometr bo'lib u chorva mollarining oyog'iga o'rnatiladi va ularning kun davomidagi faolligi bo'yicha ma'lumotlarni to'plash uchun xizmat qiladi. *Pedometr* - bu akselerometr asosida ishlaydigan, qadamlar sonini, bosib o'tilgan masofani va ba'zan jismoniy faoliyatning boshqa parametrlarini o'lchaydigan elektron qurilma hisoblanadi. Hozirgi vaqtda pedometrlar boshqa sohalar bilan bir qatorda chorva mollarida ularning umumiy jismoniy faoliyatini kuzatish uchun keng qo'llanila boshlandi.

Pedometrlardan olingan ma'lumotlardan foydalangan holda chorva mollarining oqsoqligini aniqlash modelini qurishda dastlab kiruvchi va chiquvchi parametrlarni aniqlashtirib olish lozim. Tadqiqot davomida pedometrlar yordamida quyidagicha ma'lumotlar yig'ib borildi:

- $S_c$  – har soatdagi qadamlar soni;