

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова»
Институт математики, физики и компьютерных наук
Кафедра информационных систем и методов искусственного интеллекта

Утверждена на заседании
Ученого совета ИМФКН
«___» _____ 20__ г.
Протокол №__

Рабочая программа дисциплины

Информационные системы искусственного интеллекта в медицине

Направление подготовки
02.04.03 Математическое обеспечение и администрирование
информационных систем

Квалификация
Магистр

Форма обучения
Очная

Улан-Удэ

2025

Пояснительная записка

Цели освоения дисциплины

Дисциплина «Информационные системы искусственного интеллекта в медицине» направлена на формирование у обучающегося навыков работы в информационных медицинских системах, визуализации и интерпретации результатов анализа данных медицинской аналитики. Это позволит врачу будущего обладать дополнительными знаниями, умениями, позволяющие использовать современные технологии в лечебном деле, проводить автоматизированный анализ данных о пациентах и интерпретировать результаты анализа.

1. Освоение принципов проектирования и применения информационных систем и их разновидностей для медицинских нужд.
 2. Изучение базовых понятий информационных систем, их разновидностей и различий.
 3. Освоение способов хранения данных в информационных системах и проектирования хранилищ, в том числе в зависимости от типа проводимого анализа данных.
 4. Изучение различных видов анализа данных в информационноаналитических и освоение критериев выбора разных типов анализа для различных задач.
- Изучение способов отображения информации, а также результатов проведенного анализа.

Место дисциплины в структуре образовательной программы

Дисциплина изучается в 2 семестре и является обязательной, входит в вариативную часть блока Б1 ОП по направлению подготовки 02.04.03 «Математическое обеспечение и администрирование информационных систем» (магистратура). Основывается частично на курсах: "Методы и технологии глубокого обучения", "Обработка и анализ биомедицинских сигналов".

Планируемые результаты обучения по дисциплине и индикаторы достижения компетенций.

В результате освоения дисциплины студент должен:

Знать:

Знание особенностей обработки разнородной информации от внешних источников и принципы проектирования информационного хранилища в зависимости от задач, решаемых ИС.

Уметь:

Умение разрабатывать компоненты информационных и информационноаналитических систем с учетом форматов исходных данных, требуемых результатов анализа и представления данных.

Владеть:

Демонстрирует навыки тестирования и интеграции разработанных информационных систем.

Планируемые результаты освоения образовательной программы:

- | | |
|--------|---|
| ПК-1 | Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования |
| ПК-1.1 | Выбирает и разрабатывает программные компоненты систем |

- искусственного интеллекта
- ПК-1.2 Проводит экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта
- ПК-4 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях
- ПК-4.3 Исследует и анализирует развитие новых направлений и перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное самообучение и развитие адаптивности алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)
- ПК-5 Способен организовать исполнение работ в проектах малого и среднего уровня сложности в области ИТ, в том числе при разработке инновационных биотехнических систем и технологий медицинского и физиологического назначения
- ПК-5.1 Проводит научные исследования в области создания инновационных биотехнических систем и технологий
- ПК-5.2 Выполняет проектирование инновационных биотехнических систем и технологий
- ПК-5.3 Осуществляет подготовку производства инновационных биотехнических систем

Объем дисциплины в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 3 зачетные единицы, 108 часа.

№	Название разделов дисциплины	Лекция	Лабораторная работа	Самостоятельная работа
Семестр 2		16	32	60
1	Основной модуль	16	32	60

Тематическое планирование курса

Темы

Основной модуль

Семестр 2

Введение

Лекция. 4(0) ч. Задачи курса. Цель и назначение информационных систем. Цели разработки медицинских информационных систем.

Понятие информационных систем. Виды информационных систем

Лекция. 2(0) ч. Обзор существующих видов информационных систем.

Информационноаналитические системы. Информационнокоммуникационные системы.

Особенности их применения. Выбор информационной системы в зависимости от решаемой задачи. Обзор медицинских информационных систем.

Лабораторная работа. 6(0) ч. Обзор медицинских информационных систем

Самостоятельная работа. 20(0) ч. ИДЗ № 1 «Обработка разнородных наборов медицинских данных»

Информационное пространство как среда анализа

Лекция. 2(0) ч. Понятие информационного пространства. Структура информационного пространства. Элементы структуры информационного пространства. Понятие показателя. Пространственная интерпретация понятия показатель. Содержание экономических показателей. Системы показателей.

Лабораторная работа. 6(0) ч. Обработка медицинских данных

Технологии извлечения, преобразования и загрузки данных

Лекция. 2(0) ч. Источники данных в медицинских системах. Понятие о гибкой архитектуре данных. Сбор, повышение качества данных и приведение их в единую структуру

Лабораторная работа. 6(0) ч. Разработка информационного хранилища для медицинской ИС.

Самостоятельная работа. 20(0) ч. ИДЗ № 2 «Проектирование информационного хранилища»

Концепции организации хранения данных

Лекция. 2(0) ч. Концепция информационного хранилища. Концепция централизованного хранилища данных. Концепция распределенного хранилища данных. Концепция автономных витрин данных. Концепция единого интегрированного хранилища и многих витрин данных.

Лабораторная работа. 6(0) ч. Технологии анализа данных для выполнения различных видов анализа медицинских данных.

Искусственный интеллект в медицинских системах. Технологии оперативного и интеллектуального анализа данных

Лекция. 2(0) ч. Понятие искусственного интеллекта. Подходы к выполнению анализа средствами информационных технологий (ИТанализа) в медицинских системах. Классификация ИТанализа по режиму и темпу. Требования, предъявляемые к OLAPсистемам. Задачи и содержание оперативного (OLAP) анализа. Типы многомерных OLAPсистем. Интеллектуальный анализ данных Data mining. Содержание понятия знания. Классификация видов знаний. Задачи Data mining. Специфические методы и области применения data mining.

Лабораторная работа. 6(0) ч. Анализ больших данных

Самостоятельная работа. 20(0) ч. ИДЗ № 3 «Анализ медицинских данных»

Заключение

Лекция. 2(0) ч. Подведение итогов курса

Лабораторная работа. 2(0) ч. Подведение итогов курса

БРС

Семестр	Контрольные точки	Баллы
2	Текущий контроль в разделе «Основной модуль»	
	Практическое задание	20
	Практическое задание	20
	Практическое задание	20
2	Зачет	
	Теоретические вопросы	40

Итого за семестр 2: 100

Учебно-методическое и информационное обеспечение учебного процесса

Образовательные технологии (в том числе на занятиях, проводимых в интерактивных формах).

Периодически используется технология проблемного обучения. Студентам даются сырые статистические данные из некоторой прикладной области (психология, социология, медицина, экономика и т.д.). Преподаватель формулирует задачу с точки зрения предметной области. Студенту необходимо правильно формализовать задачу и выбрать соответствующий метод анализа, затем решить ее с помощью специализированного программного пакета. Полученный результат необходимо интерпретировать с точки зрения предметной области.

Лекционные и лабораторные занятия проводятся с использованием презентаций.

- [Методические рекомендации к практическим работам.pdf](#)

Режим доступа: https://my.bsu.ru/content/file/5/52/522/1095669_metodicheskie-rekomendacii-k-prakticheskim-rabotam.pdf

Учебно-методические материалы, в том числе методические указания для обучающихся по освоению дисциплины

Методические указания к лекционным занятиям.

1. Если во время лекции все же не совсем разобрались в отдельных моментах раскрываемой темы, рекомендуется в индивидуальном порядке уточнить непонятные разделы у преподавателя во время лекции (поднять руку и задать вопрос), либо после нее.

Скромность - качество замечательное, но в отдельных случаях 4
быть скромным просто неразумно.

2. Для того, чтобы составлять качественные конспекты лекций, важно понять, что конспект – не дословно записанная речь преподавателя. Преподаватель вообще не обязан диктовать текст лекции под запись, он ведет согласно плану. Таким образом, в течение лекции студент тратит большую часть времени на восприятие информации, меньшую его часть – на ее запись.

3. Для повышения эффективности конспектирования материала рекомендуется воспользоваться следующими рекомендациями: 1) Убирайте только середину слова, а не середину и окончание (например, удачный «эф-ть», не удачный «эф.»). 2) В процессе лекции пишите часть слова, затем в тексте оставляйте место для второй его части, а на перерыве или после занятий (пока не забыли, о чем шла речь) вписывайте оставшуюся часть слова.

3) Заменяйте длинные русские слова короткими иностранными, например, несколько – some, выигрывать – win, использовать – use, экономический – economic и т.д.

Методические указания к лабораторным (практическим) занятиям.

1. В ходе лабораторных занятий обучающиеся фактически впервые сталкиваются с самостоятельной практической деятельностью в конкретной области – содействует становлению студентов как будущих специалистов. Поэтому, необходимо студенту проявить здесь особое усердие и получить ощутимый результат.

2. Результаты выполнения лабораторных (практических) работ нужно оформить в виде отчета. Как правило, отчет состоит из 3-х частей: план отчета (общая структура задания); расчетные формулы, блок-схема алгоритма, принципиальная часть программного кода, применяемые методы и средства (библиотеки, модули, структуры данных, службы, шаблоны классов, математические методы ит.п.), авторский проект решения задачи; выводы.

3. Перед сдачей лабораторных работ (практических заданий) необходимо повторить теоретический материал для более глубокого понимания и грамотного комментирования выполненной работы преподавателю.

Методические указания к самостоятельной работе студента.

1. Выполняйте внеаудиторное задание в день его получения, а накануне занятия повторите его.

2. Для успешного выполнения задания создайте условия, которые отвечают требованиям гигиены умственного труда: удобное место, достаточное освещение, тишина, перерывы, необходимое оборудование.

3. Начинайте выполнять задание с его осмысления: определите цель, содержание, степень новизны, уровень усвоения, объем, сроки, этапы и приемы выполнения. Спланируйте и соблюдайте затем последовательность действий. Познакомьтесь с алгоритмом и эталоном выполнения задания.

4. Изучите вначале теоретическую основу задания (закон, правило, первоисточник и др.), затем принимайтесь за практическую работу.

5. Старайтесь выполнять задание самостоятельно, применяя знания и умения, усвоенные ранее.

6. Определите свой оптимальный ритм и режим работы.

7. Помните, что следование рекомендациям научной организации учебного труда экономит время, способствует достижению наилучших результатов.

Оценочные средства

По данной дисциплине разработаны оценочные средства, критерии их оценивания, а также методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций (в приложении).

Список литературы

Перечень основной и дополнительной литературы, необходимой для освоения дисциплины.

Основная

1. [Интеллектуальные системы и технологии](#): Учебник и практикум для вузов/Станкевич Л. А.. —Москва: Юрайт, 2022. —397 с.
Режим доступа: <https://urait.ru/bcode/489694>
2. [Интеллектуальные системы. Теория хранения и поиска информации](#): Учебник для вузов/Гасанов Э. Э., Кудрявцев В. Б.. —Москва: Юрайт, 2022. —271 с.
Режим доступа: <https://urait.ru/bcode/491100>
3. [Интеллектуальные системы](#): Учебное пособие для вузов/Иванов В. М. ; под науч. ред. Сесекина А.Н.. —Москва: Юрайт, 2022. —91 с.
Режим доступа: <https://urait.ru/bcode/492094>
4. [Интеллектуальные системы](#): Учебник и практикум для вузов/Бессмертный И. А., Нугуманова А. Б., Платонов А. В.. —Москва: Юрайт, 2022. —243 с.
Режим доступа: <https://urait.ru/bcode/490020>
5. [Интеллектуальные системы](#): Учебник и практикум для вузов/Кудрявцев В. Б., Гасанов Э. Э., Подколзин А. С.. —Москва: Юрайт, 2022. —165 с.
Режим доступа: <https://urait.ru/bcode/491107>
6. [ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ](#): Учебник и практикум/Бессмертный И.А., Нугуманова А.Б., Платонов А.В.. —М.: Издательство Юрайт, 2016. —243 с.
Режим доступа: <http://www.biblio-online.ru/book/C9C644CA-F8C0-4CE8-BD2D-26AB852CCAF8>

Дополнительная

1. Искусственный интеллект : междисциплинарный подход/[А. Ю. Алексеев [и др.] ; под ред. Д. И. Дубровского, В. А. Лекторского]. —М.: [ИИИнтелЛ], 2006. —444 с.
2. [Занимательный искусственный интеллект. Манга](#)/Сакамото М.. —Москва: ДМК Пресс, 2022. —188 с.
Режим доступа: <https://e.lanbook.com/book/314885>
3. [Интеллектуальные системы. Теория и приложения](#)/ООО "Интеллектуальные системы". —Москва. —2411-4448.
Режим доступа: https://elibrary.ru/title_about.asp?id=54513
4. [ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ](#): Учебник и практикум/Кудрявцев В.Б., Гасанов Э.Э., Подколзин А.С.. —М.: Издательство Юрайт, 2016. —219 с.
Режим доступа: <http://www.biblio-online.ru/book/1DAA117E-A40C-4F22-B6EA-642C255D29CB>
5. [ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ. ТЕОРИЯ ХРАНЕНИЯ И ПОИСКА ИНФОРМАЦИИ](#): Учебник/Гасанов Э.Э., Кудрявцев В.Б.. —М.: Издательство Юрайт, 2016. —289 с.
Режим доступа: <http://www.biblio-online.ru/book/2771E75A-5B2D-4E2D-BD2B-B13DFB2916EB>
6. [ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ](#): Учебник и практикум/Станкевич Л.А.. —М.: Издательство Юрайт, 2016. —397 с.
Режим доступа: <http://www.biblio-online.ru/book/750B2832-EE3F-4E58-9F52-037C8E916959>

Перечень ресурсов информационно-коммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотечная система «Медицинская библиотека «MEDLIB.RU» (www.medlib.ru)
2. Электронная медицинская библиотека «Консультант врача» (www.rosmedlib.ru)
3. Полнотекстовая база данных «ClinicalKey» (www.clinicalkey.com)
5. HTS The Biomedical & Life Sciences Collection – 2400 аудиовизуальных презентаций (www.hstalks.com)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Личный кабинет преподавателя или студента БГУ <https://my.bsu.ru/>

GitHub (веб-сервис для хостинга IT-проектов и их совместной разработки):
<https://github.com/>

Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Кабинет для лекционных занятий с мультимедийным оборудованием.

Компьютерный класс с мультимедийным оборудованием.

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова»
Институт математики, физики и компьютерных наук
Кафедра информационных систем и методов искусственного интеллекта

Фонд оценочных средств по учебной дисциплине

Информационные системы искусственного интеллекта в медицине

Направление подготовки
02.04.03 Математическое обеспечение и администрирование
информационных систем

Квалификация
Магистр

Форма обучения
Очная

Улан-Удэ
2025

**Паспорт фонда оценочных средств (ФОС) по дисциплине
«Информационные системы искусственного интеллекта в медицине»**

Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы:

- ПК-1 Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования
- ПК-1.1 Выбирает и разрабатывает программные компоненты систем искусственного интеллекта
- ПК-1.2 Проводит экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта
- ПК-4 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях
- ПК-4.3 Исследует и анализирует развитие новых направлений и перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное самообучение и развитие адаптивности алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)
- ПК-5 Способен организовать исполнение работ в проектах малого и среднего уровня сложности в области ИТ, в том числе при разработке инновационных биотехнических систем и технологий медицинского и физиологического назначения
- ПК-5.1 Проводит научные исследования в области создания инновационных биотехнических систем и технологий
- ПК-5.2 Выполняет проектирование инновационных биотехнических систем и технологий
- ПК-5.3 Осуществляет подготовку производства инновационных биотехнических систем

Этапы формирования компетенции

Семестр	Вид контроля	Оценочные средства
2 семестр	Текущий	Лабораторные работы
	Итоговый (зачет)	Теоретические вопросы

Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Фонд оценочных средств сформирован на основе ключевых принципов оценивания:

- валидность: объекты оценки должны соответствовать поставленным целям обучения;
- определенность: оценочные средства должны быть понятны каждому обучающемуся;
- однозначность: одинаковость оценки качества оценочного средства;
- надежность: использование единообразных показателей и критериев для оценивания достижений.

Описание показателей и критериев оценивания компетенций, а также шкал оценивания

Показатели оценивания компетенций	Уровень сформированности компетенций	Шкала оценивания	Официальный цифровой эквивалент оценки
Знать: особенности обработки разнородной информации от внешних источников и принципы проектирования информационного хранилища в зависимости от задач, решаемых ИС. Уметь: разрабатывать компоненты информационных и информационно-аналитических систем с учетом форматов исходных данных, требуемых результатов анализа и представления данных. Владеть: навыками тестирования и интеграции разработанных информационных систем.	Высокий	85 – 100 баллов	5 (отлично)
	Базовый	70 – 84 баллов	4 (хорошо)
	Пороговый	60-69 баллов	3 (удовлетворительно)

Балльно-рейтинговая система

Для текущего и итогового контроля качества обучения студентов и магистрантов применяется балльно-рейтинговая система, разработанная в соответствии с «Положением об организации учебного процесса с применением кредитно-модульной системы обучения», утвержденным Учебно-методическим советом ФГБОУ ВО «Бурятский государственный университет» от 20 февраля 2012 г. Целью БРС является определение уровня успешности освоения (завершения изучения) обучающимися учебных дисциплин (модулей, циклов) через балльные оценки и рейтинги качества сформированных знаний, умений, профессиональных компетенций, накапливаемые в соответствии с измеряемыми в зачетных единицах трудоемкостями каждого цикла (модуля, дисциплины) и основной образовательной программы в целом.

1. Общая максимальная сумма баллов, которую студент может набрать по дисциплине в течение семестра – 100 баллов: 60 баллов текущий контроль и 40 баллов итоговый контроль (зачет).

2. Минимальная сумма баллов, при которой студент допускается к зачету (итоговому контролю), равна 20 баллам.

3. Минимальная сумма баллов, при которой студент получает положительную итоговую оценку (зачет) по дисциплине равна 60 баллам (60% от 100 баллов).

4. Максимальная оценка за выполнение одной лабораторной работы – 20 баллов.

Связь между четырехбалльной и стобалльной системами оценки качества обучения студентов

Оценка	Рейтинговые баллы
Отлично	85-100
Хорошо	70-84
Удовлетворительно	60-69

ПРИМЕРЫ ЛАБОРАТОРНЫХ РАБОТ (ТЕКУЩИЙ КОНТРОЛЬ)

Лабораторная работа № 1.

Тема: «Обработка разнородных наборов медицинских данных для построения интегрированной модели прогнозирования»

1. Цель работы:

Сформировать единый набор признаков (feature set) из разнородных медицинских данных (структурированные числовые данные, категориальные данные и текстовые выписки) и построить на его основе модель прогнозирования, оценив преимущества интеграции данных по сравнению с использованием только структурированной информации.

2. Задачи:

1. Загрузить и провести разведочный анализ разнородных данных.
2. Выполнить предобработку и feature engineering для каждого типа данных:
 - a. Числовые данные
 - b. Категориальные данные
 - c. Текстовые данные (медицинские выписки)
3. Объединить обработанные данные в единую матрицу "объекты-признаки".
4. Построить и оценить модели машинного обучения на основе:
 - a. Только структурированных данных (числовые + категориальные)
 - b. Только текстовых данных
 - c. Объединенного набора данных
5. Сравнить эффективность моделей и сделать выводы.

3. Используемое программное обеспечение и данные:

- **Язык программирования:** Python 3
- **Библиотеки:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, NLTK/spaCy (для NLP), WordCloud.
- **Данные:** Для реалистичности можно сгенерировать синтетический датасет, содержащий:
 - **Числовые признаки:** Возраст, ИМТ, уровень глюкозы, артериальное давление.
 - **Категориальные признаки:** Пол, наличие вредных привычек (курение, алкоголь), семейный анамнез.
 - **Текстовые признаки:** Краткая выписка из истории болезни с симптомами (например: "Пациент жалуется на кашель, повышенную температуру и слабость").
 - **Целевая переменная:** Диагноз (например, "Здоров", "Грипп", "ОРВИ") или риск госпитализации (0/1).

4. Теоретическое введение:

Разнородные данные (Multimodal Data): Данные, поступающие из различных источников и имеющие разную природу и структуру.

Проблемы обработки: Несогласованность форматов, пропуски, разные шкалы измерений, проблема "проклятия размерности".

Feature Engineering: Процесс создания и отбора признаков, наиболее информативных для задачи прогнозирования.

Обработка текста (NLP): Методы векторизации текста: "Мешок слов" (Bag-of-Words), TF-IDF, N-граммы.

Ход работы:

Этап 1: Загрузка и разведочный анализ данных (EDA)

1. Загрузите данные из CSV-файлов (или сгенерируйте синтетические данные).

2. Проведите первичный анализ:

- `df.info()`, `df.describe()`
- Визуализация распределения числовых признаков (гистограммы, `boxplots`).
- Анализ распределения категориальных признаков (столбчатые диаграммы).
- Визуализация наиболее частых слов в текстовых выписках с помощью облака слов (`WordCloud`).
- Проверка на пропущенные значения и выбросы.

Этап 2: Предобработка и инжиниринг признаков

2.1. Обработка числовых данных:

- Заполнение пропущенных значений (медианой или средним).
- Стандартизация данных (`StandardScaler`) для приведения признаков к единому масштабу.

2.2. Обработка категориальных данных:

- Преобразование строковых категорий в числовые с помощью `One-Hot Encoding` или `Label Encoding`.

2.3. Обработка текстовых данных (самый сложный и творческий этап):

- **Очистка текста:** Приведение к нижнему регистру, удаление пунктуации, стоп-слов (например, "и", "в", "на"), цифр.
- **Лемматизация/Стемминг:** Приведение слов к их нормальной форме (например, "жаловался" -> "жаловаться").
- **Векторизация текста:** Преобразование очищенного текста в числовой вектор с помощью метода **TF-IDF**. Можно экспериментировать с количеством признаков (`max_features`) и использовать n-граммы (например, биграммы "высокая температура").

2.4. Объединение данных:

- Используйте `np.hstack()` или `pd.concat()` для объединения трех полученных матриц (числовой, категориальной и текстовой) в одну общую матрицу признаков `X_combined`.

Этап 3: Построение и оценка моделей

1. Разделите данные на обучающую и тестовую выборки (`train_test_split`).
2. Обучите модели (например, **Логистическая регрессия** или **Случайный лес**) на трех различных наборах данных:
 - **Модель 1:** Только числовые + категориальные признаки (`X_structured`).
 - **Модель 2:** Только текстовые признаки (`X_text`).
 - **Модель 3:** Объединенный набор всех признаков (`X_combined`).
3. Для каждой модели сделайте прогноз на тестовой выборке и рассчитайте метрики:
 - **Accuracy** (Точность)
 - **Precision, Recall, F1-Score** (для каждого класса)
 - **Macro-average F1** (если задача многоклассовая)
 - **AUC-ROC** (если задача бинарной классификации)

Этап 4: Сравнительный анализ и выводы

- Создайте сводную таблицу с метриками для всех трех моделей.
- Постройте общую столбчатую диаграмму для сравнения Accuracy или F1-Score моделей.
- Визуализируйте матрицы ошибок для каждой модели.

Содержание отчета:

1. Титульный лист.
2. Цель и задачи работы.
3. Теоретические сведения. Краткое описание методов обработки каждого типа данных.

4. **Описание набора данных.**
5. **Блок-схема алгоритма** всего процесса (от загрузки данных до оценки модели).
6. **Результаты выполнения программы:**
 - Графики EDA (распределения, облако слов).
 - Примеры обработанного текста (до/после очистки).
 - Сводная таблица с метриками и сравнительные графики.
 - Матрицы ошибок.
7. **Анализ полученных результатов:**
 - Какая модель показала наилучший результат?
 - Насколько значительный прирост в качестве дало объединение разнородных данных?
 - Какие типы признаков (числовые, категориальные, текстовые) оказались наиболее информативными?
 - С какими основными трудностями столкнулись при обработке текста?
8. **Выводы.** Обобщение проделанной работы и полученного опыта.

Лабораторная работа № 2

Тема: «Проектирование и прототипирование информационного хранилища для анализа медицинских данных»

1. Цель работы:

Спроектировать концептуальную, логическую и физическую модель тематически ориентированного, интегрированного, неизменяемого и привязанного ко времени хранилища данных (Data Warehouse) для поддержки аналитических задач в медицинской системе искусственного интеллекта.

2. Задачи:

1. Определить бизнес-требования и предметную область.
2. Выделить основные бизнес-процессы и разработать концептуальную модель хранилища (схему «звезда» или «снежинка»).
3. Спроектировать логическую модель: определить таблицы фактов и измерений, их атрибуты и типы данных.
4. Реализовать физическую модель в СУБД (например, PostgreSQL).
5. Наполнить хранилище тестовыми данными и выполнить базовые аналитические запросы.

3. Используемое программное обеспечение:

Средства проектирования: Draw.io, Lucidchart, MS Visio или pgModeler.

Система управления базами данных (СУБД): PostgreSQL / MySQL.

Средство для ETL-процесса: Python с библиотеками (Pandas, SQLAlchemy) или средствами СУБД (INSERT...SELECT).

Ход работы:

Этап 1: Определение предметной области и бизнес-требований

- Выбор предметной области: Например, «Анализ эффективности лечения пациентов с сердечно-сосудистыми заболеваниями» или «Мониторинг лабораторных исследований и постановки диагнозов».
- Формулировка бизнес-вопросов, на которые должно отвечать хранилище:
 - Какова динамика количества пациентов по месяцам и отделениям?
 - Каков средний срок лечения для конкретного диагноза?
 - Какие лекарства назначаются наиболее часто при определенном заболевании?
 - Какова загрузка врачей различного профиля?

Этап 2: Проектирование концептуальной модели (Схема «Звезда»)

Выбираем бизнес-процесс — например, «Лабораторный анализ».

- Таблица фактов (Fact_Lab_Analysis):
 - Содержит *измеряемые* данные (метрики).
 - Ключевые столбцы: analysis_id (PK), patient_key (FK), doctor_key (FK), time_key (FK), analysis_type_key (FK).
 - Показатели (Measures): result_value (числовой результат анализа), cost (стоимость анализа).
- Таблицы измерений (Dimensions):
 - Dim_Patient: patient_key (PK), patient_id, gender, birth_date, blood_type.
 - Dim_Doctor: doctor_key (PK), doctor_id, full_name, specialization, department.
 - Dim_Time: time_key (PK), full_date, day, month, quarter, year.
(Критически важная таблица для любого хранилища)
 - Dim_Analysis_Type: analysis_type_key (PK), analysis_name, unit_of_measure, normal_range_min, normal_range_max.

Этап 3: Разработка логической и физической модели

- Логическая модель: Детализируем каждую таблицу, определяем типы данных (INT, VARCHAR, DATE, DECIMAL), первичные и внешние ключи.
- Физическая модель:
 - Создаем новую базу данных в СУБД: CREATE DATABASE medical_dw;
 - Написываем и выполняем DDL-скрипты (Data Definition Language) для создания таблиц.

Этап 4: Наполнение хранилища данными (ETL-процесс)

1. Создаем тестовые данные для таблиц измерений (например, 5-10 записей в каждую).
2. Пишем скрипт на Python или SQL для заполнения таблицы фактов, связывая данные из измерений.

Этап 5: Выполнение аналитических запросов

Написать и выполнить SQL-запросы, отвечающие на бизнес-вопросы из Этапа 1.

- Запрос 1: «Общее количество анализов по специализациям врачей за последний квартал»
- Запрос 2: «Средний результат анализа глюкозы по месяцам»

Содержание отчета:

1. Титульный лист.
2. Цель и задачи работы.
3. Теоретическая часть: Краткое описание концепции хранилища данных, схем «звезда»/«снежинка», ETL-процесса.
4. Описание предметной области и бизнес-требований.
5. Концептуальная модель: Готовая схема «звезда» (диаграмма).
6. Логическая и физическая модель:
 - ER-диаграмма или подробное описание таблиц.
 - DDL-скрипты создания всех таблиц.
7. Наполнение данными:
 - Примеры тестовых данных (скриншоты или CSV-файлы).
 - Код ETL-скрипта (Python или SQL).
8. Аналитические запросы:
 - Формулировка бизнес-вопроса.
 - Текст SQL-запроса.
 - Результат выполнения (скриншот или таблица).
9. Выводы:

- Какие преимущества дает хранилище данных перед операционной БД для аналитики?
- С какими сложностями столкнулись при проектировании?
- Как спроектированное хранилище может быть использовано для обучения моделей ИИ? (Например, как источник очищенных и интегрированных данных для построения признаков).

Лабораторная работа № 3.

Тема: «Анализ медицинских данных с использованием методов статистики и машинного обучения»

1. Цель работы:

Освоить полный цикл анализа медицинских данных: от разведочного анализа и проверки статистических гипотез до построения прогностических моделей машинного обучения для решения клинических задач.

2. Задачи:

1. Провести первичный разведочный анализ данных (EDA).
2. Выполнить предобработку и очистку данных.
3. Проверить статистические гипотезы.
4. Построить и оценить прогностические модели.
5. Интерпретировать результаты с медицинской точки зрения.

3. Используемое программное обеспечение:

- **Язык программирования:** Python
- **Библиотеки:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, SciPy
- **Данные:** Pima Indians Diabetes Dataset / Heart Disease Dataset / COVID-19 Radiography Dataset

Ход работы:

Этап 1: Первичный анализ данных (EDA)

1. **Загрузка данных:**
2. **Анализ структуры данных:**
 - Размерность данных
 - Типы признаков
 - Пропущенные значения
 - Дубликаты
3. **Статистический анализ:**
4. **Визуализация распределений:**
 - Гистограммы для числовых признаков
 - Boxplot для выявления выбросов
 - Countplot для категориальных признаков

Этап 2: Предобработка данных

1. **Обработка пропущенных значений:**
 - Удаление или импутация (среднее, медиана)
2. **Обработка выбросов:**

- Метод межквартильного размаха (IQR)
- Winsorization
- 3. **Кодирование категориальных признаков**
- 4. **Нормализация данных**
- Этап 3: Статистический анализ
- 1. **Проверка нормальности распределения:**
 - Тест Шапиро-Уилка
- 2. **Сравнение групп:**
 - t-тест для нормальных распределений
 - U-тест Манна-Уитни для ненормальных
- 3. **Корреляционный анализ:**
 - Матрица корреляций
 - Тепловая карта
- Этап 4: Построение прогностических моделей
- 1. **Подготовка данных**
- 2. **Обучение моделей**
- 3. **Кросс-валидация**
- Этап 5: Оценка моделей
- 1. **Метрики качества**
- 2. **ROC-кривая**
- 3. **Важность признаков**

Содержание отчета:

1. **Титульный лист**
2. **Цель и задачи работы**
3. **Описание набора данных**
4. **Методы и инструменты анализа**
5. **Результаты:**
 - Результаты EDA (графики, таблицы)
 - Результаты статистического анализа
 - Сравнение моделей по метрикам
 - Матрицы ошибок и ROC-кривые
 - Важность признаков
6. **Анализ полученных результатов:**
 - Какие факторы наиболее значимы для прогноза?
 - Какая модель показала наилучший результат и почему?
 - Какие статистически значимые различия обнаружены между группами?
 - Как результаты могут быть применены в клинической практике?
7. **Выводы**

Критерии оценки:

- Полнота проведения EDA
- Корректность статистического анализа
- Качество предобработки данных
- Разнообразие и адекватность выбранных моделей
- Глубина интерпретации результатов
- Оформление отчета и кода

ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ НА ЗАЧЕТ (ИТОГОВЫЙ КОНТРОЛЬ)

Блок 1: Основные понятия и архитектура систем

1. Дайте определение информационной системы искусственного интеллекта (ИС ИИ) в медицине. Назовите основные компоненты.
2. Классификация медицинских ИС ИИ по целям применения (диагностические, прогностические, терапевтические, административные).
3. В чем заключаются основные преимущества и ограничения использования ИИ в медицине?
4. Опишите жизненный цикл разработки медицинской ИС ИИ.
5. Что такое "интерпретируемость" (interpretability) и "объяснимость" (explainability) моделей ИИ и почему они критически важны в медицине?

Блок 2: Данные в медицинских ИС ИИ

6. Особенности и виды медицинских данных (структурированные, неструктурированные, полуструктурированные). Приведите примеры.
7. Опишите основные этапы предобработки медицинских данных перед их использованием в моделях ИИ.
8. В чем заключаются проблемы и методы работы с пропущенными данными (missing data) в медицинских наборах?
9. Что такое "несбалансированные выборки" (class imbalance)? Какие методы борьбы с ними вы знаете?
10. Требования к обеспечению конфиденциальности и безопасности медицинских данных (ФЗ-152, GDPR, HIPAA).

Блок 3: Основные технологии и методы ИИ

11. Классификация задач машинного обучения: обучение с учителем, без учителя, с подкреплением. Примеры в медицине.
12. Основные этапы построения прогностической модели машинного обучения.
13. Методы оценки качества моделей классификации и регрессии в медицинских задачах (Accuracy, Precision, Recall, F1-Score, AUC-ROC). Какую метрику выбрать, если важно не пропустить больных пациентов?
14. Принцип работы и применение деревьев решений и случайного леса (Random Forest) в медицине.
15. Области применения нейронных сетей в медицине. Опишите архитектуру сверточной нейронной сети (CNN) для анализа изображений.
16. Что такое обработка естественного языка (NLP)? Примеры использования NLP для анализа медицинских текстов (истории болезни, выписки).
17. В чем разница между моделью машинного обучения и экспертной системой? Приведите примеры экспертных систем в медицине.

Блок 4: Прикладные аспекты и конкретные применения

18. ИС ИИ для диагностики по медицинским изображениям (рентген, КТ, МРТ, гистология). Примеры успешных применений.
19. Прогностическое моделирование (predictive analytics) в медицине: для прогнозирования рисков заболеваний, рецидивов, продолжительности госпитализации.
20. Применение ИИ в персонализированной и прецизионной медицине.

21. Роль ИИ в ускорении разработки лекарственных средств (drug discovery).
22. Системы поддержки врачебных решений (Clinical Decision Support System - CDSS): архитектура и принципы работы.
23. Использование ИИ для мониторинга пациентов и анализа данных в реальном времени (телемедицина, IoT-устройства).

Критерии оценки ответа на теоретический вопрос

Критерий	Требования	Баллы
Знание и понимание теоретического материала	Рассматриваемые понятия определяются четко и полно, приводятся соответствующие примеры	10
	Используемые понятия строго соответствуют теме	8
Анализ и оценка информации	Объясняются альтернативные взгляды на рассматриваемую проблему	8
	Дается личная оценка проблеме	6
Построение суждений	Изложение ясное и четкое	4
	Приводятся различные точки зрения и их личная оценка (примеры применения, плюсы и минусы)	4
	Итого	40