

Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ КемГУ
Дата и время: 2025-09-24 00:00:00
471086fad29a3b30e244c728abc3661ab35c9d50210dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Кузбасский гуманитарно-педагогический институт
федерального государственного бюджетного образовательного учреждения высшего образования
«Кемеровский государственный университет»
Факультет физической культуры, естествознания и природопользования

ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ
дисциплины
Машинное обучение в здравоохранении

Специальность
30.05.03 Медицинская кибернетика

Направленность (профиль)
«Медицинские информационные системы»

Программа специалитета

Новокузнецк, 2025

Оглавление

1. Общие положения	3
1.1 Общие сведения о фонде оценочных материалов дисциплины	3
1.2. Порядок формирования и оценивания выполнения теста.....	3
2 ФОМ дисциплины «Машинное обучение в здравоохранении»	3
2.1 Назначение ФОМ дисциплины	3
2.2 Результаты освоения дисциплины	3
3. Диагностические задания по дисциплине «Машинное обучение в здравоохранении»	4

1. Общие положения

1.1 Общие сведения о фонде оценочных материалов дисциплины

Фонд оценочных материалов дисциплины (ФОМ) содержит не менее 40 заданий закрытого и открытого типов, в том числе не менее 20-ти заданий закрытого типа и 20-ти заданий открытого типа (таблица 1) для формирования не менее 2-х вариантов тестов, предъявляемых студентам учебной группы для диагностического тестирования.

Таблица 1 – Структура ФОМ дисциплины и минимальное количество заданий по типам и видам

Типы и виды заданий ФОМ дисциплины	Минимальное количество заданий в ФОМ
Задания закрытого типа:	20
1. задания с выбором одного или нескольких ответов;	
2. задания на сопоставление;	
3. задания на установление правильной последовательности.	
Задания открытого типа:	20
1. задания на дополнение;	10
2. задания с развернутым ответом.	10
ИТОГО ЗАДАНИЙ	40

Для многосеместровой дисциплины общее количество заданий в ФОМ и количество заданий по типам и видам может превышать минимально установленное в п. 1.1.

1.2. Порядок формирования и оценивания выполнения теста

Комплект заданий (тест) для проверки результатов освоения дисциплины формируется из заданий ФОМ дисциплины. Максимальное количество заданий в тесте – 20 (10 заданий закрытого типа, 10 заданий открытого типа).

На выполнение теста из 20-ти заданий обучающемуся на контрольном мероприятии выделяется 2 академических часа.

Максимальное количество баллов, которое можно получить за выполнение теста – 20 баллов. Оценка, которую может получить студент в зависимости от количества баллов, набранных за выполнение всех заданий теста, в отношении к максимальном возможному, представлена в таблице 2.

Таблица 2 – Шкала оценивания уровня освоения дисциплины

Сумма набранных баллов	Уровни освоения	Экзамен		Зачет
		Оценка	Буквенный эквивалент	Буквенный эквивалент
17,2- 20	Продвинутый	5	отлично	Зачтено
13,2- 17,1	Повышенный	4	хорошо	
10 – 13,1	Пороговый	3	удовлетворительно	
0 – 9,9	Первый	2	неудовлетворительно	Не зачтено

2 ФОМ дисциплины «Машинное обучение в здравоохранении»

2.1 Назначение ФОМ дисциплины

ФОМ дисциплины Машинное обучение в здравоохранении предназначен для контроля результатов освоения дисциплины в ходе промежуточной аттестации в форме зачета и экзамена по итогам полного изучения учебного материала всех семестров.

ФОМ может использоваться в текущей аттестации в ходе изучения дисциплины и в семестровой промежуточной аттестации.

2.2 Результаты освоения дисциплины

В результате освоения дисциплины Машинное обучение в здравоохранении у обучающихся формируются профессиональные компетенции: ПК-3, ПК-4.

3. Диагностические задания по дисциплине «Машинное обучение в здравоохранении»

Диагностические задания	Количество заданий
	Ключи к заданиям (эталонные ответы), критерии оценки
ЗАДАНИЯ ЗАКРЫТОГО ТИПА	20
Задания с выбором одного или нескольких ответов	8
<p>Задание 1. Прочитайте текст, выберите все правильные ответы</p> <p>Какие из следующих задач относятся к задачам обучения с учителем?</p> <ol style="list-style-type: none"> 1) Предсказание количества одноразовых шприцов, которое понадобится в следующем месяце в процедурной, исходя из всей истории работы процедурного кабинета 2) Кластеризация пациентов по их показателям здоровья 3) Предсказание специальности врача, которым было написано назначение лечения, при условии, что у вас есть доступ ко всевозможным текстам назначений различных врачей 4) Предсказание, кем был написан текст выписки из больницы: мужчиной или женщиной. В качестве входных данных есть 100 тысяч текстов выписок из этой больницы без какой-либо дополнительной информации 5) Прогноз количества пациентов, которые обратятся с симптомами гриппа в поликлиники города в текущем месяце, на основании исторических данных о пациентах поликлиник за последние 3 года. 	<p><i>Эталонный ответ – 135</i></p> <p><i>Критерии оценки</i> 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.</p>
<p>Задание 2. Прочитайте текст, выберите правильный ответ</p> <p>Что является основной целью применения линейной регрессии в медицинских исследованиях?</p> <ol style="list-style-type: none"> 1) Определение точной зависимости между переменными 2) Прогнозирование значения одной непрерывной переменной на основе других 3) Классификация пациентов на группы по признакам болезни 4) Поиск оптимальных значений дозировок лекарств 5) Подбор правильной схемы лечения 	<p><i>Эталонный ответ – 2</i></p> <p><i>Критерии оценки</i> правильный ответ – 1 б.</p>
<p>Задание 3. Прочитайте текст, выберите все правильные ответы</p> <p>Среди функций выберите те, которые могли бы стать ядрами в методе опорных векторов</p> <ol style="list-style-type: none"> 1) $K(\mathbf{a}, \mathbf{b}) = \sin(\mathbf{a}^T \mathbf{b})$ 	<p><i>Эталонный ответ – 245</i></p> <p><i>Критерии оценки</i> 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.</p>

<p>2) $K(\mathbf{a}, \mathbf{b}) = (\mathbf{a}, \mathbf{b})^2$</p> <p>3) $K(\mathbf{a}, \mathbf{b}) = (\mathbf{a} - \mathbf{b})^2$</p> <p>4) $K(\mathbf{a}, \mathbf{b}) = \tanh((\mathbf{a}, \mathbf{b}) + 3)$</p> <p>5) $K(\mathbf{a}, \mathbf{b}) = \exp(-2 \cdot \ \mathbf{a} - \mathbf{b}\ ^2)$</p>	
<p>Задание 4. Прочитайте текст, выберите правильный ответ</p> <p>Предположим, мы построили линейную регрессию для прогнозирования веса тела пациента исходя из роста. Что означают стандартные остатки в данном анализе?</p> <p>1) Различия между фактическими значениями веса и теми, которые были спрогнозированы моделью</p> <p>2) Среднее значение веса всех исследуемых пациентов</p> <p>3) Сумма весов пациентов</p> <p>4) Стандартизированные отклонения возраста пациентов</p>	<p>Эталонный ответ – 1</p> <p>Критерии оценки правильный ответ – 1 б.</p>
<p>Задание 5. Прочитайте текст, выберите правильный ответ</p> <p>Критерий, используемый для выбора лучшей модели при решении задачи регрессии среди нескольких вариантов, называется...</p> <p>1) F-тест</p> <p>2) Тест Манна-Уитни</p> <p>3) t-критерий Стьюдента</p> <p>4) Индекс детерминации (R^2)</p> <p>5) Полнота</p>	<p>Эталонный ответ – 4</p> <p>Критерии оценки 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.</p>
<p>Задание 6. Прочитайте текст, выберите все правильные ответы</p> <p>Решающие деревья применяются в медицине для решения следующих задач:</p> <p>1) Диагностика заболеваний на основе симптомов</p> <p>2) Выявление взаимосвязей между симптомами и результатами анализов</p> <p>3) Прогнозирование сроков выздоровления после операций</p> <p>4) Оценка эффективности медикаментозного лечения</p> <p>5) Планирование расписания приема врачей</p>	<p>Эталонный ответ – 1234</p> <p>Критерии оценки 4 правильных ответа – 1 б. 3 правильных ответа – 0,75 б. 2 правильных ответа – 0,5 б. 1 правильный ответ – 0,25 б.</p>
<p>Задание 7. Прочитайте текст, выберите все правильные ответы</p> <p>Недостатки метода решающих деревьев в медицине проявляются следующим образом:</p> <p>1) Склонность к переобучению на специфичных примерах</p> <p>2) Невозможность моделировать нелинейные связи между признаками</p> <p>3) Чувствительность к небольшим изменениям структуры дерева</p> <p>4) Низкая устойчивость к пропущенным данным</p> <p>5) Трудности в обработке сильно зашумленных данных</p>	<p>Эталонный ответ – 135</p> <p>Критерии оценки 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.</p>

Задание 8. Прочитайте текст, выберите все правильные ответы Методы кластеризации активно используются в медицине для: 1) Разработки персонализированных методов профилактики хронических заболеваний 2) Определения типов инфекционных возбудителей 3) Прогнозирования потребности в ресурсах здравоохранения 4) Установления диагноза конкретного пациента 5) Улучшения качества планирования медицинского обслуживания регионов				Эталонный ответ – 135 Критерии оценки 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.	
Задания на сопоставление				7	
Задание 9. Прочитайте текст и установите соответствие Соотнесите конкретную задачу с типом задачи машинного обучения				Эталонный ответ – А2Б4В3Г1 Критерии оценки 4 правильных ответа – 1 б. 3 правильных ответа – 0,75 б. 2 правильных ответа – 0,5 б. 1 правильный ответ – 0,25 б.	
прикладная задача		тип задачи машинного обучения			
А	Возможность определить, является ли комментарий в соцсети токсичным	1	кластеризация		
Б	Предсказание оценки студента за экзамен. Оценка измеряется по пятибалльной шкале	2	бинарная классификация		
В	Предсказание стоимости расходов на обслуживание клиента по полису дополнительного медицинского страхования	3	регрессия		
Г	Возможность определить, сколько различных профилей пользователей пользуется приложением по доставке еды. Профиль — комбинация характеристик типичного пользователя	4	многоклассовая классификация		
Задание 10. Прочитайте текст и установите соответствие Соотнесите задачу машинного обучения с метрикой				Эталонный ответ – А3Б1В2 Критерии оценки 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.	
задача		метрика			
А	классифицировать новостные заголовки и отнести их к какой-то одной категории. Текстов под разные категории примерно одинаковое количество.	1	recall (полнота)		
Б	медицинская задача бинарной классификации некоторой болезни	2	F1-мера		
В	задача кредитного скоринга: определяем, выдать кредит заёмщику	3	ассигасу (доля правильно классифициро-		

	или нет. Заёмщик может быть хорошим, платить кредит вовремя, или плохим — допускать просрочку по оплате.		ванных объектов)	
Задание 11. Прочитайте текст и установите соответствие Соотнесите формулы расстояний и их названия				Эталонный ответ – А1Б3В2 Критерии оценки 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.
Формула расстояния		Название расстояния		
А	$L^2 = \sqrt{\sum_{m=1}^n \left(x_m^{(i)} - x_m^{(j)}\right)^2}$	1	Евклидово расстояние	
Б	$L^1 = \sum_{m=1}^n \left x_m^{(i)} - x_m^{(j)}\right $	2	Расстояние Чебышева	
В	$L^\infty = \max \left x_m^{(i)} - x_m^{(j)}\right $	3	Манхэттенское расстояние	
Задание 12. Прочитайте текст и установите соответствие Соотнесите метод машинного обучения и функционал ошибки, используемый для оптимизации				Эталонный ответ – А3Б2В1 Критерии оценки 3 правильных ответа – 1 б. 2 правильных ответа – 0,66 б. 1 правильный ответ – 0,33 б.
Метод машинного обучения		Функционал ошибки		
А	Линейная регрессия	1	$L = \frac{\ \mathbf{w}\ ^2}{2C} + \sum_{i=1}^n \left(1 - M_i(\mathbf{w})\right)_+$	
Б	Логистическая регрессия	2	$L = -\frac{1}{n} \sum_{i=1}^n \left(y_i \cdot \log(y'_i) + (1 - y_i) \cdot \log(1 - y'_i)\right)$	
В	Метод опорных векторов	3	$L = \frac{\sum_{i=1}^m \left(y_i - y'_i\right)^2}{m}$	
Задание 13. Прочитайте текст и установите соответствие Соотнесите термины алгоритма k-средних и их определения				Эталонный ответ – А2Б3В1Г4 Критерии оценки 4 правильных ответа – 1 б. 3 правильных ответа – 0,75 б. 2 правильных ответа – 0,5 б. 1 правильный ответ – 0,25 б.
Термин		Определение		
А	Центроид	1	Расстояние от объекта до ближайшего центра кластера	
Б	Объект	2	Точка, характеризующая среднее положение всех элементов своего кластера	
В	Радиус	3	Совокупность признаков отдельного пациента или образца	
Г	К (число кластеров)	4	Определяет количество заранее заданных групп, на которое будут разбиты объекты	
Задание 14. Прочитайте текст и установите соответствие				Эталонный ответ – А2Б4В1Г3

Соотнесите типы ансамблей и методов объединения решений				<i>Критерии оценки</i> 4 правильных ответа – 1 б. 3 правильных ответа – 0,75 б. 2 правильный ответ – 0,5 б. 1 правильный ответ – 0,25 б.	
Тип ансамбля		Сценарий применения			
А	Voting Classifier	1	Решения объединяются градиентным способом, стремясь минимизировать ошибку		
Б	AdaBoost	2	Решение принимают путём простого или взвешенного голосования		
В	Gradient Boosting	3	Модели строятся независимо друг от друга и голосование осуществляется простым большинством		
Г	Bagging	4	Весовые коэффициенты классов определяются динамически, завися от ошибок предыдущего классификатора		
Задание 15. <i>Прочитайте текст и установите соответствие</i> Соотнесите преимущества ансамблевых методов с областью их применения в медицине					<i>Эталонный ответ</i> – АЗБ2В4Г1 <i>Критерии оценки</i> 4 правильных ответа – 1 б. 3 правильных ответа – 0,75 б. 2 правильный ответ – 0,5 б. 1 правильный ответ – 0,25 б.
Преимущество		Применение			
А	Повышение устойчивости к шуму данных	1	Даже при небольшом числе пациентов ансамбль позволяет получать хорошие результаты		
Б	Получение лучших показателей точности	2	Повышенная способность различать раковые клетки и здоровые ткани		
В	Снижение эффекта переобучения	3	Уменьшается зависимость от аномалий и выбросов в данных, что важно при анализе данных КТ или МРТ		
Г	Эффективность в условиях малого объема данных	4	Более точные прогнозы исходов операций и диагнозов		
Задания на установление правильной последовательности					5
Задание 16. <i>Прочитайте текст и установите последовательность</i> Расположите шаги алгоритма кросс-валидации в правильном порядке: 1) Повторение процесса для каждой части 2) Оценка модели на оставшейся части 3) Разделение данных на k частей 4) Обучение модели на k-1 частей 5) Усреднение результатов оценки					<i>Эталонный ответ</i> – 34215 <i>Критерии оценки</i> Верная последовательность – 1 б.
Задание 17. <i>Прочитайте текст и установите последовательность</i> Установите правильную последовательность обучения и настройки линейной регрессии для прогнозирования давления крови:					

1) Установка гиперпараметра регуляризации. 2) Исследование корреляционных связей между признаками. 3) Загрузка и первичный анализ набора данных. 4) Финальная калибровка модели и получение оценочных показателей качества. 5) Отбор значимых признаков для модели.	
Задание 18. <i>Прочитайте текст и установите последовательность</i> Установите правильный порядок этапов при обучении нейронной сети: 1) Вычисление функции потерь 2) Обратное распространение (backpropagation) 3) Инициализация весов 4) Прямое распространение (forward propagation) 5) Обновление весов	<i>Эталонный ответ</i> – 31425 <i>Критерии оценки</i> Верная последовательность – 1 б.
Задание 19. <i>Прочитайте текст и установите последовательность</i> Расположите шаги обучения модели машинного обучения в правильном порядке: 1) Оценка модели на тестовой выборке 2) Выбор алгоритма машинного обучения 3) Обучение модели на обучающей выборке 4) Подбор гиперпараметров 5) Выбор метрики для оценки модели	<i>Эталонный ответ</i> – 25431 <i>Критерии оценки</i> Верная последовательность – 1 б.
Задание 20. <i>Прочитайте текст и установите последовательность</i> Расположите процедуры предобработки данных перед использованием метода k-средних для сегментации пациентов в правильном порядке: 1) Устранение пропущенных значений или замена средним / медианой. 2) Удаление дубликатов записей. 3) Запуск метода k-means с предварительно установленным числом кластеров. 4) Масштабирование признаков (например, стандартизацией Z-нормализацией).	<i>Эталонный ответ</i> – 2143 <i>Критерии оценки</i> Верная последовательность – 1 б.
ЗАДАНИЯ ОТКРЫТОГО ТИПА	20
Задания на дополнение	10
Задание 21. <i>Прочитайте текст и дополните ответ</i> При построении модели можно получить явление _____, при котором построенная модель хорошо объясняет примеры из обучающей выборки, но относительно плохо работает на примерах, не участвовавших в обучении.	<i>Эталонный ответ</i> – переобучение <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв будет верным)</i>
Задание 22. <i>Прочитайте текст и дополните ответ</i> Отсутствие значимого различия между значениями метрик на обучающей и тестовой выборке будет свидетельствовать то, что модель _____.	<i>Эталонный ответ</i> – не переобучилась <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв будет верным)</i>

Задание 23. <i>Прочитайте текст и дополните ответ</i> _____ это связь признаков модели через строгую функциональную зависимость.	<i>Эталонный ответ</i> – мультиколлинеарность <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв будет верным)</i>
Задание 24. <i>Прочитайте текст и дополните ответ</i> Разности между реальными значениями и значениями, предсказанными регрессионной моделью называют _____ регрессии (residuals)	<i>Эталонный ответ</i> – остатки <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв будет верным)</i>
Задание 25. <i>Прочитайте текст и дополните ответ</i> Чтобы построенная модель линейной регрессии была корректной, остатки должны быть _____. Это свойство, обозначающее постоянство дисперсии некоторой последовательности случайных величин	<i>Эталонный ответ</i> – гомоскедастичными <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв, не искажающее смысл ответа, будет верным)</i>
Задание 26. <i>Прочитайте текст и дополните ответ</i> _____ в машинном обучении вносит в обучающую функцию систему «штрафов» за слишком большие коэффициенты и помогает избавляться от ненужных свойств (шумов).	<i>Эталонный ответ</i> – регуляризация <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв будет верным)</i>
Задание 27. <i>Прочитайте текст и дополните ответ</i> Гипотеза _____ в задачах классификации это предположение о том, что схожие объекты гораздо чаще лежат в одном классе, чем в разных.	<i>Эталонный ответ</i> – компактности <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв, не искажающее смысл ответа, будет верным)</i>
Задание 28. <i>Прочитайте текст и дополните ответ</i> Гипотеза _____ в задачах регрессии предположение о том, что близким объектам соответствуют близкие ответы.	<i>Эталонный ответ</i> – непрерывности <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв, не искажающее смысл ответа, будет верным)</i>
Задание 29. <i>Прочитайте текст и дополните ответ</i> Когда классы в наборе данных не сбалансированы, модель машинного обучения становится более предвзятой к мажорному классу, что приводит к плохим результатам обработки данных _____ класса	<i>Эталонный ответ</i> – минорного <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв, не искажающее смысл ответа, будет верным)</i>
Задание 30. <i>Прочитайте текст и дополните ответ</i> Чтобы снизить вероятность возникновения проблемы затухания градиентов при построении свёрточной нейронной сети, нужно после свёрточных слоев выбрать функцию активации _____.	<i>Эталонный ответ</i> – ReLU <i>Критерии оценки</i> Верный ответ – 1 б. <i>(любое количество строчных и прописных букв, не искажающее смысл ответа, будет верным)</i>
Задания с развернутым ответом	10
Задание 31. <i>Прочитайте текст и запишите развернутый от-</i>	<i>Эталонный ответ</i> 1) Диагностика заболеваний (рак, диабет,

<p><i>вет</i></p> <p>Перечислите не менее 5 основных направлений применения машинного обучения в современной медицине.</p>	<p>сердечно-сосудистые патологии).</p> <p>2) Прогнозирование течения болезней и рисков осложнений.</p> <p>3) Поддержка принятия врачебных решений (разработка рекомендательных систем).</p> <p>4) Индивидуализация подходов к лечению (персонализация медицины).</p> <p>5) Управление ресурсами здравоохранения (логистика, распределение персонала).</p> <p>6) Анализ большого массива медицинских данных (Big Data) для открытия новых закономерностей.</p> <p>7) Инновационные технологии скрининга (например, выявление COVID-19 по рентгенограммам лёгких).</p> <p>8) Мониторинг здоровья пациентов удалённо (носимая электроника, телемедицина).</p> <p>9) Улучшение процедур реабилитации и восстановления функций организма.</p> <p>10) Научные исследования и разработка новых медикаментов.</p> <p><i>Критерии оценки:</i></p> <p>0 б. – нет верных ответов.</p> <p>0,2 б. – 1 верный ответ</p> <p>0,4 б. – 2 верных ответа</p> <p>0,6 б. – 3 верных ответа</p> <p>0,8 б. – 4 верных ответа</p> <p>1 б. – 5 верных ответов</p>
<p>Задание 32.</p> <p><i>Прочитайте текст и запишите развернутый ответ</i></p> <p>Перечислите не менее 5 инструментов и библиотек Python, популярных для анализа медицинских данных с помощью машинного обучения. Укажите для решения каких задач они используются</p>	<p><i>Эталонный ответ</i></p> <p>1) Pandas (для загрузки, очистки и преобразования данных).</p> <p>2) NumPy (математические вычисления над массивами данных).</p> <p>3) Scikit-Learn (машинное обучение, классификация, регрессия, кластеризация).</p> <p>4) TensorFlow/Keras (для глубоких нейронных сетей и анализа изображений).</p> <p>5) PyTorch (гибкое и мощное фреймворк для нейронных сетей).</p> <p>6) Matplotlib/Seaborn (визуализация данных).</p> <p>7) Statsmodels (эконометрические и статистические модели).</p> <p>8) Shapely/Geopandas (работа с географическими данными).</p> <p>9) SciPy (расширенные научные расчеты и статистика).</p> <p>10) Imbalanced-learn (решение проблем несбалансированности классов).</p> <p><i>Критерии оценки:</i></p> <p>0 б. – 0 правильных ответов,</p> <p>0,25 б. – 1 правильный ответ,</p> <p>0,5 б. – 2 правильных ответа,</p> <p>0,75 б. – 3 правильных ответа,</p> <p>1 б. – 4 правильных ответа</p>
<p>Задание 33.</p> <p><i>Прочитайте текст и запишите развернутый ответ</i></p>	<p><i>Эталонный ответ</i></p> <p>1) Репрезентативность данных (представленность всей популяции).</p>

<p>Перечислите не менее 4 ключевых требований к наборам данных для успешного обучения моделей машинного обучения в медицине.</p>	<ol style="list-style-type: none"> 2) Качество и чистота данных. 3) Достаточно большой объём данных для адекватного обучения модели. 4) Однородность представления данных (единый формат, единая шкала измерений). 5) Минимальное количество пропусков и отсутствие систематических пробелов. 6) Правильный баланс классов. 7) Независимость и случайность отбора объектов. 8) Надежность измерительных приборов и методик сбора данных. 9) Постоянная верификация и обновление данных. <p><i>Критерии оценки:</i> 0 б. – 0 правильных ответов, 0,25 б. – 1 правильный ответ, 0,5 б. – 2 правильных ответа, 0,75 б. – 3 правильных ответа, 1 б. – 4 правильных ответа</p>
<p>Задание 34. <i>Прочитайте текст и запишите развернутый ответ</i></p> <p>Перечислите не менее 4 этапов предобработки данных для построения модели машинного обучения.</p>	<p><i>Эталонный ответ</i></p> <ol style="list-style-type: none"> 1) Нормализация числовых данных 2) Удаление выбросов 3) Обработка аномальных значений 4) Кодирование категориальных признаков 5) Разделение данных на обучающую и тестовую выборки 6) Заполнение пропущенных значений <p><i>Критерии оценки:</i> 0 б. – нет верных ответов. 0,25 б. – 1 верный ответ 0,5 б. – 2 верных ответа 0,75 б. – 3 верных ответа 1 б. – 4 верных ответа</p>
<p>Задание 35. <i>Прочитайте текст и запишите развернутый ответ</i></p> <p>Каковы основные стратегии работы с несбалансированными данными? Дайте развернутое описание стратегий</p>	<p><i>Эталонный ответ</i></p> <ol style="list-style-type: none"> 1) Изменить алгоритмы машинного обучения, учитывая асимметричное распределение классов, придав мажорному и минорному классам разные веса. А именно сильнее штрафовать модель за неправильную классификацию минорного класса, установив ему более высокий вес, чем мажорному. 2) Дублировать записи минорного класса до тех пор, пока размеры классов не будут сбалансированы (алгоритм upsampling). 3) Удалить случайным образом экземпляры мажорного класса (алгоритм downsampling) 4) Сгенерировать новые экземпляры минорного класса из существующих, но не идентичные им (алгоритм SMOTE). <p><i>Критерии оценки:</i> 0 б. – нет верных ответов. 0,25 б. – 1 верный ответ 0,5 б. – 2 верных ответа 0,75 б. – 3 верных ответа</p>

	1 б. – 4 верных ответа
<p>Задание 36. Прочитайте текст и запишите развернутый ответ</p> <p>Рассмотрим пример: человек сдает анализ на СПИД. В качестве нулевой гипотезы возьмем: «Человек здоров».</p> <p>1) Обоснуйте выбор нулевой гипотеза 2) Сформулируйте альтернативную гипотезу 3) Опишите ошибку первого рода 4) Опишите ошибку второго рода 5) Определите какая из двух ошибок более опасна в этом случае</p>	<p><i>Эталонный ответ</i></p> <p>1) Выбираем нулевой гипотезой «Человек здоров», так как большинство людей здоровы), 2) Альтернативная гипотеза - человек болен. 3) Ошибкой первого рода будет, если здоровому человеку анализ покажет, что он болен. 4) Ошибкой второго рода будет, если больному человеку анализ покажет, что он здоров. 5) Ошибки второго рода намного опаснее, потому что больному человеку не будет оказана соответствующая медицинская помощь и его жизнь будет поставлено под угрозу.</p> <p><i>Критерии оценки:</i> 0 б. – нет верных ответов. 0,2 б. – 1 верный ответ 0,4 б. – 2 верных ответа 0,6 б. – 3 верных ответа 0,8 б. – 4 верных ответа 1 б. – 5 верных ответов</p>
<p>Задание 37. Прочитайте текст и запишите развернутый ответ</p> <p>Существует несколько подходов для устранения эффекта мультиколлинеарности. Перечислите не менее 3.</p>	<p><i>Эталонный ответ</i></p> <p>1) Исключить из модели одну из дублирующих друг друга переменных. 2) Преобразовать дублирующие признаки в меньшее число новых признаков. 3) Преобразовать признаки с помощью получения смещённых оценок.</p> <p><i>Критерии оценки:</i> 0 б. – нет верных ответов. 0,33 б. – 1 верный ответ 0,66 б. – 2 верных ответа 1 б. – 3 верных ответа</p>
<p>Задание 38. Прочитайте текст и запишите развернутый ответ</p> <p>Допустим, перед вами стоит задача регрессии по предсказанию числа пациентов приемного отделения за день. Какие можно использовать признаки? Перечислите не менее 4.</p>	<p><i>Эталонный ответ</i></p> <p>1) набор симптомов, 2) пол пациентов, 3) число пациентов за определённый период в прошлом, 4) день недели, 5) выходной день или праздничный, 6) время года, 7) уровень эпидемиологической обстановки.</p> <p><i>Критерии оценки:</i> 0 б. – нет верных ответов. 0,25 б. – 1 верный ответ 0,5 б. – 2 верных ответа 0,75 б. – 3 верных ответа 1 б. – 4 верных ответа</p>
<p>Задание 39. Прочитайте текст и запишите развернутый ответ</p> <p>Определите метод линейной регрессии. Для чего:</p> <p>1) Напишите уравнение линейной регрессии в об-</p>	<p><i>Эталонный ответ</i></p> <p>1) $y = f(w, x) = w_0 + w_1x_1 + \dots + w_kx_k$ 2) y — целевая переменная, (x_1, \dots, x_k) — вектор признаков, w_0, w_1, \dots, w_k — параметры модели: w_1, \dots, w_k — вектор весов, w_0 — свободный коэффициент или сдвиг (bias).</p>

<p>щем виде</p> <p>2) Определите его составляющие.</p> <p>3) Укажите название метода, используемого для построения функционала для минимизации</p> <p>4) Запишите формулу функционала потерь.</p>	<p>3) При решении задачи минимизируется квадратичная функция потерь. Метод называется методом наименьших квадратов, или МНК</p> $Loss = \sum_{i=1}^n (y_i - f(x_i))^2 \rightarrow \min$ <p>4)</p> <p><i>Критерии оценки:</i></p> <p>0 б. – нет верных ответов.</p> <p>0,25 б. – 1 верный ответ</p> <p>0,5 б. – 2 верных ответа</p> <p>0,75 б. – 3 верных ответа</p> <p>1 б. – 4 верных ответа</p>
<p>Задание 40.</p> <p><i>Прочитайте текст и запишите развернутый ответ</i></p> <p>Перечислите не менее 4 критериев оценивания качества моделей машинного обучения в медицине.</p>	<p><i>Эталонный ответ</i></p> <ol style="list-style-type: none"> 1) Accurasy (доля верно классифицированных объектов). 2) Precision (доля верно классифицированных положительных случаев). 3) Recall (чувствительность, доля выявленных истинно позитивных случаев). 4) F1 мера (средневзвешенное гармоническое значение precision и recall). 5) Area Under Curve (ROC-AUC) (качество классификации на кривой ROC). 6) Mean Absolute Error (MAE) (средняя абсолютная ошибка). 7) Root Mean Square Error (RMSE) (квадратичная средняя абсолютная ошибка). 8) Specificity (специфичность, процент ложноположительных случаев). 9) Матрица ошибки (confusion matrix) для наглядного отображения ошибок. 10) Время отклика и вычислительная нагрузка модели. <p><i>Критерии оценки:</i></p> <p>0 б. – нет верных ответов.</p> <p>0,25 б. – 1 верный ответ</p> <p>0,5 б. – 2 верных ответа</p> <p>0,75 б. – 3 верных ответа</p> <p>1 б. – 4 верных ответа</p>
ИТОГО:	40

Составитель:

Решетникова Е.В., доцент кафедры математики, физики и математического моделирования.
Ф.И.О. должность, наименование кафедры