

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ ВЛАДИМИРА ДАЛЯ»

Краснодонский факультет инженерии и менеджмента (филиал)  
Кафедра информационных технологий и транспорта



УТВЕРЖДАЮ:

Директор

Панайотов К.К.

(подпись)

«14» марта 2025 года

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

по учебной дисциплине

**Машинное обучение и анализ данных**

(наименование учебной дисциплины, практики)

**09.04.01 Информатика и вычислительная техника**

(код и наименование направления подготовки (специальности))

**«Интеллектуальные системы**

**в производственно-транспортных комплексах»**

наименование профиля подготовки (специальности, магистерской программы); при отсутствии ставится прочерк)

Разработчик(разработчики):

доцент

(подпись)

Бихдрикер А. С.

ФОС рассмотрен и одобрен на заседании кафедры информационных технологий и транспорта от «26» февраля 2025 г., протокол № 7

Заведующий кафедрой

информационных

технологий и транспорта

(подпись)

Верительник Е. А.

Краснодон 2025

**Комплект оценочных материалов по дисциплине  
«Машинное обучение и анализ данных»**

**Задания закрытого типа**

***Задания закрытого типа на выбор правильного ответа***

1. *Выберите один правильный ответ.*

Выберите верное определение для термина переобучение (overfitting) в контексте машинного обучения?

- А) Ситуация, когда модель слишком обобщена.
- Б) Ситуация, когда модель слишком хорошо подстроилась под тренировочные данные.
- В) Ситуация, когда модель слишком проста для данных.
- Г) Ситуация, когда модель слишком сложная для данных.

Правильный ответ: Б.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

2. *Выберите один правильный ответ.*

Метод прогнозирования категориальных значений в машинном обучении?

- А) Регрессия.
- Б) Кластерный анализ.
- В) Классификация.
- Г) Обучение с учителем.

Правильный ответ: В.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

3. *Выберите один правильный ответ.*

Выберите верное определение для кросс-валидации в машинном обучении?

- А) Метод обучения без учителя.
- Б) Метод оценки производительности модели.
- В) Метод оптимизации модели.
- Г) Метод обучения с учителем.

Правильный ответ: Б.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

4. *Выберите один правильный ответ.*

Выберите виды обучения, которые существуют в машинном обучении?

- А) Обучение с учителем и без учителя.
- Б) Обучение с учителем, без учителя и с подкреплением.
- В) Обучение с учителем и с подкреплением.

Правильный ответ: Б.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

5. Выберите один правильный ответ.

Выберите наиболее часто используемую метрику для оценки качества модели классификации?

А) F1-мера.

Б) Среднеквадратичная ошибка (MSE).

В) Коэффициент детерминации

Г) Средняя абсолютная ошибка (MAE).

Правильный ответ: А.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

### ***Задания закрытого типа на установление соответствия***

1. Типы задач машинного обучения и примеры приложений. Установите соответствие типу задач машинного обучения и приложению. Каждому элементу левого столбца соответствует только один элемент правого столбца:

| Тип                         | Приложение                                           |
|-----------------------------|------------------------------------------------------|
| 1) Классификация            | А) Определение спама в электронной почте.            |
| 2) Регрессия                | Б) Прогнозирование стоимости жилья.                  |
| 3) Кластеризация            | В) Сегментация аудитории для маркетинговых кампаний. |
| 4) Обучение с подкреплением | Г) Обучение игрового агента в компьютерных играх.    |
|                             | Д) Генерация изображений по текстовым запросам.      |

Правильный ответ: 1-А, 2-Б, 3-В, 4-Г

Компетенции (индикаторы): ПК-2 (ПК-2.2).

2. Методы оценки качества моделей машинного обучения и их описание. Установите соответствие методу оценки качества моделей и его описанию. Каждому элементу левого столбца соответствует только один элемент правого столбца:

| Метод                | Описание                                                                                                                            |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| 1) 1 Кросс-валидация | А) График, отображающий степень корректности классификации в зависимости от выбранного порогового значения.                         |
| 2) 2 Матрица ошибок  | Б) Разбиение данных на $n$ частей и последовательное использование каждой части в качестве теста, а оставшихся в качестве обучения. |
| 3) 3 MSE / RMSE      | В) Таблица, используемая для оценки точности предсказаний бинарной                                                                  |

классификации при несбалансированных классах.

- Г) Среднеквадратичная ошибка / корень из среднеквадратичной ошибки для оценки точности регрессионных моделей.

Правильный ответ: 1-Б, 2-В, 3-Г

Компетенции (индикаторы): ПК-2 (ПК-2.2).

3. Основные этапы построения модели машинного обучения и описание. Установите соответствие этапу построения модели и его описанию. Каждому элементу левого столбца соответствует только один элемент правого столбца:

| Этап                                          | Описание                                                                                                                                                                                     |
|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1) Подготовка данных                          | А) Используется для улучшения точности и обобщающей способности модели. Включает в себя анализ ошибок, настройку параметров модели, а также использование регуляризации и ансамблей моделей. |
| 2) Выбор метода и создание модели             | Б) Этап, на котором проверяется способность модели к обобщению на новые данные, не использовавшиеся в процессе обучения.                                                                     |
| 3) Оценка и улучшение качества модели         | В) Этап, на котором осуществляется разметка данных.                                                                                                                                          |
| 4) Оценка качества модели на тестовой выборке | Г) Этап, на котором осуществляется сбор, очистка, преобразование данных, а также разделение на обучающую и тестовую выборки.                                                                 |
| 5)                                            | Д) На этом этапе определяется, какой алгоритм машинного обучения будет использоваться, настраиваются параметры модели, и происходит ее обучение на обучающей выборке.                        |

Правильный ответ: 1-Г, 2-Д, 3-А, 4-Б

Компетенции (индикаторы): ПК-2 (ПК-2.2).

4. Примеры алгоритмов машинного обучения и их описание. Установите соответствие примерам алгоритмов машинного обучения их описанию. Каждому элементу левого столбца соответствует только один элемент правого столбца:

| Алгоритм                         | Описание                                                                                        |
|----------------------------------|-------------------------------------------------------------------------------------------------|
| 1) SVM (Support Vector Machines) | А) Алгоритм классификации, основанный на построении гиперплоскостей в многомерном пространстве. |
| 2) K-Means                       | Б) Алгоритм, способный строить деревья                                                          |

- |    |                                         |                                                                                                                                                                                            |
|----|-----------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                         | решений на основе случайных подвыборок признаков и наблюдений, затем комбинируя результаты.                                                                                                |
| 3) | Random Forest                           | В) Алгоритм кластеризации, который разбивает набор данных на заранее заданное количество кластеров.                                                                                        |
| 4) | Наивный<br>Байесовский<br>классификатор | Г) Алгоритм регрессии, основанный на линейной зависимости между признаками и целевым значением.<br>Алгоритм классификации на основе вероятностного подхода, основанного на теореме Байеса. |

Правильный ответ: 1-А, 2-В, 3-Б, 4-Д

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*5. Типы ансамблей моделей машинного обучения и описание. Установите соответствие типу ансамблей моделей их описанию. Каждому элементу левого столбца соответствует только один элемент правого столбца:*

- | Тип                                | Описание                                                                                                                                                                                                                                                                                         |
|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1) Баггинг (Bootstrap Aggregating) | А) Метод, при котором несколько моделей голосуют за окончательный прогноз, взвешенное или равномерное количество голосов разных моделей может быть использовано.                                                                                                                                 |
| 2) Бустинг (Boosting)              | Б) Способ построения последовательности слабых моделей, каждая следующая обучается на данных, на которых предыдущая допускала ошибки.                                                                                                                                                            |
| 3) Стекинг (Stacking)              | В) Метод построения нескольких моделей на основе случайных подвыборок обучающей выборки, агрегирование результатов для получения окончательного прогноза.<br>Г) Метод, в котором несколько моделей обучаются независимо, а затем их результаты используются в качестве признаков в общей модели. |

Правильный ответ: 1-В, 2-Б, 3-Г

Компетенции (индикаторы): ПК-2 (ПК-2.2).

***Задания закрытого типа на установление правильной последовательности***

*1. Расположите по возрастанию сложности задачи машинного обучения. Запишите правильную последовательность букв слева направо:*

А) 1. Классификация.

Б) 2. Регрессия.

В) 3. Кластеризация.

Правильный ответ: Б, А, В

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*2. Упорядочите следующие шаги построения модели машинного обучения по порядку. Запишите правильную последовательность букв слева направо:*

А) Обучение модели.

Б) Оценка модели.

В) Предварительная обработка данных.

Правильный ответ: В, А, Б

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*3. Упорядочите виды обучения с учителем по степени сложности. Запишите правильную последовательность букв слева направо:*

А) Классификация.

Б) Регрессия.

В) Обучение с подкреплением.

Правильный ответ: Б, А, В

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*4. Расположите по порядку шаги борьбы с переобучением модели. Запишите правильную последовательность букв слева направо:*

А) Увеличение объёма обучающей выборки.

Б) Регуляризация.

В) Уменьшение сложности модели.

Правильный ответ: В, А, Б

Компетенции (индикаторы): ПК-2 (ПК-2.2).

## **Задания открытого типа**

### ***Задания открытого типа на дополнение***

*1. Вставьте пропущенное слово.*

Регрессия, метод опорных векторов (SVM), наивный байесовский классификатор и др. — это модели машинного обучения для решения задачи \_\_\_\_\_.

Правильный ответ: прогнозирования

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*2. Вставьте пропущенное слово.*

\_\_\_\_\_ - модель, используемая для анализа связи между числовыми переменными с использованием линейной функции.

Правильный ответ: Линейная регрессия  
Компетенции (индикаторы): ПК-2 (ПК-2.2).

3. *Вставьте пропущенное слово.*

Для построения рекомендательных систем используется алгоритм \_\_\_\_\_ фильтрации.

Правильный ответ: коллаборативной.  
Компетенции (индикаторы): ПК-2 (ПК-2.2).

4. *Вставьте пропущенное слово.*

Кластеризация, снижение размерности данных, поиск аномалий и др. - это алгоритмы в которых используется обучение \_\_\_\_\_.

Правильный ответ: обучение без учителя.  
Компетенции (индикаторы): ПК-2 (ПК-2.2).

### ***Задания открытого типа с кратким свободным ответом***

1. *Дайте ответ на вопрос.*

Дайте определение понятию «Байесовский классификатор».

Правильный ответ: Статистический метод машинного обучения, основанный на теореме Байеса.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

2. *Дайте ответ на вопрос.*

Как называются данные, которые имеют категориальное значение и не поддаются измерению по числовой шкале?

Правильный ответ: Категориальные данные / Качественные данные.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

3. *Дайте ответ на вопрос.*

Как называется метод уменьшения размерности, основанный на поиске главных компонент данных?

Правильный ответ: РСА / Метод главных компонент / Анализ главных компонент.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

4. *Дайте ответ на вопрос.*

Как называется процесс разделения данных на группы на основе их сходства?

Правильный ответ: Кластеризация / Группировка / Сегментация.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

### ***Задания открытого типа с развернутым ответом***

*1. Дайте развернутый ответ на вопрос:*

Опишите, что такое выбор признаков (Feature Selection) в контексте машинного обучения и анализа данных. Объясните, почему выбор признаков является важным этапом, какие преимущества он предоставляет, и какие существуют основные методы выбора признаков.

Время выполнения: 20 мин.

Ожидаемый результат:

Выбор признаков (Feature Selection) – это процесс выбора подмножества наиболее релевантных признаков из исходного набора признаков для построения модели машинного обучения. Это важный этап, так как использование всех доступных признаков не всегда приводит к лучшим результатам.

Преимущества выбора признаков.

1. Улучшение точности модели. Удаление нерелевантных или избыточных признаков может снизить шум и улучшить обобщающую способность модели.
2. Уменьшение времени обучения модели. Модель с меньшим количеством признаков обучается быстрее.
3. Упрощение модели. Более простую модель легче интерпретировать и понимать.
4. Снижение переобучения. Уменьшение количества признаков снижает риск переобучения модели на обучающихся данных.

Основные методы выбора признаков.

1. Фильтры. Используют статистические показатели для оценки релевантности признаков (например, коэффициент корреляции, критерий хи-квадрат).
2. Обертки. Оценивают качество подмножеств признаков, обучая и оценивая модель на каждом подмножестве (например, прямой и обратный отбор).
3. Встроенные методы. Выполняют выбор признаков в процессе обучения модели (например, L1-регуляризация в линейных моделях, важность признаков в деревьях решений).

Критерии оценивания: ответ должен содержать описание преимуществ и методов минимум одним предложением.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*2. Дайте развернутый ответ на вопрос:*

Опишите различные методы обработки пропущенных значений в наборах данных. Объясните, какие факторы следует учитывать при выборе метода обработки пропущенных значений.

Время выполнения: 20 мин.

Ожидаемый результат:

Пропущенные значения – это обычное явление в наборах данных. Их наличие может негативно влиять на результаты анализа и обучения моделей



машинного обучения. Существует несколько методов обработки пропущенных значений.

1. Удаление строк/столбцов с пропущенными значениями. Простой метод, но может привести к потере большого количества данных, если пропущенных значений много.
  2. Заполнение пропущенных значений константой. Заполнение пропущенных значений определенным значением (например, 0, -1).
  3. Заполнение пропущенных значений средним/медианой/модой. Заполнение пропущенных значений средним значением для числовых признаков, медианой - если есть выбросы, модой - для категориальных.
  4. Заполнение пропущенных значений на основе других признаков. Использование других признаков для предсказания пропущенных значений (например, с помощью регрессионной модели).
  5. Использование алгоритмов, устойчивых к пропущенным значениям. Некоторые алгоритмы машинного обучения (например, деревья решений) могут обрабатывать пропущенные значения без предварительной обработки.
- Факторы, влияющие на выбор метода.

1. Процент пропущенных значений. Если пропущенных значений очень много, удаление строк может привести к значительной потере данных.
2. Тип данных. Для числовых признаков можно использовать заполнение средним/медианой, для категориальных – модой.
3. Важность признака. Если признак важен для модели, лучше использовать более сложные методы заполнения, чем просто удаление строк.
4. Характер пропусков. Важно понимать, случайно ли пропущены значения, связаны ли они с другими признаками или с самим значением признака.

Критерии оценивания: ответ должен содержать три метода и три фактора.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*3. Дайте развернутый ответ на вопрос:*

Опишите, как оценивать качество моделей кластеризации. Объясните, какие метрики используются для оценки кластеризации, и чем они отличаются друг от друга.

Время выполнения: 20 мин.

Ожидаемый результат:

Оценка качества моделей кластеризации – важный этап, позволяющий определить, насколько хорошо алгоритм кластеризации разделил данные на группы. В отличие от задач классификации и регрессии, в задачах кластеризации часто нет «правильных» ответов, поэтому оценка качества является более сложной задачей.

Существуют два основных типа метрик для оценки кластеризации.

Внутренние метрики. Оценивают качество кластеризации на основе самих данных и результатов кластеризации, не используя внешнюю информацию (например, метки классов).

1. Индекс силуэта. Оценивает, насколько объект похож на объекты в своем кластере по сравнению с объектами в других кластерах. Значение индекса силуэта варьируется от -1 до 1, где более высокие значения указывают на лучшую кластеризацию.

2. Индекс Дэвиса-Болдина. Оценивает среднее сходство каждого кластера с его наиболее похожим кластером. Меньшие значения индекса Дэвиса-Болдина указывают на лучшую кластеризацию.

Внешние метрики. Оценивают качество кластеризации, сравнивая результаты кластеризации с известными метками классов.

1. Индекс Рэнда. Оценивает процент правильных пар объектов (объекты, которые находятся в одном кластере и имеют одинаковые метки, или объекты, которые находятся в разных кластерах и имеют разные метки).

2. Adjusted Rand index (ARI). Является нормализованной версией индекса Рэнда, учитывающей случайное соглашение между кластеризацией и метками классов.

3. Нормализованная взаимная информация. Измеряет количество информации, которой одна кластеризация делится с другой.

Выбор метрики зависит от доступности информации о данных и целей анализа.

Критерии оценивания: каждая метрика должна быть описана минимум одним предложением.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

*4. Дайте развернутый ответ на вопрос:*

Объясните, что такое ансамблевые методы машинного обучения. Опишите общую концепцию, лежащую в основе ансамблей, их преимущества и недостатки по сравнению с отдельными моделями.

Время выполнения: 20 мин.

Ожидаемый результат:

Ансамблевые методы в машинном обучении – это методы, объединяющие предсказания нескольких базовых моделей (слабых учеников) для получения более точных и устойчивых предсказаний. Основная идея заключается в том, что коллективный «разум» нескольких моделей часто превосходит производительность любой отдельной модели.

Преимущества ансамблей.

1. Повышенная точность – ансамбли обычно показывают лучшую производительность по сравнению с отдельными моделями за счет усреднения или комбинирования предсказаний.

2. Устойчивость – ансамбли менее подвержены переобучению и лучше обобщаются на новые данные.

3. Улучшенная обобщающая способность – ансамбли могут охватывать различные аспекты данных и захватывать сложные зависимости.

Недостатки ансамблей:

1. Сложность - обучение и настройка ансамблей может быть более сложной задачей, чем обучение отдельной модели.
2. Вычислительные затраты - обучение и предсказание с использованием ансамблей может занимать больше времени и ресурсов, чем для одной модели.
3. Интерпретируемость – ансамбли могут быть менее интерпретируемыми, чем отдельные модели, особенно, если они состоят из большого количества компонентов.

Ансамбли показывают лучшую производительность, потому что они уменьшают дисперсию, смещение или и то и другое. Усреднение предсказаний различных моделей уменьшает чувствительность к выбросам и шуму в данных. Объединение моделей, обученных на разных подмножествах данных или с использованием различных алгоритмов, может компенсировать недостатки отдельных моделей.

Критерии оценивания: ответ должен содержать определение, преимущества и недостатки, которые описаны минимум одним предложением.

Компетенции (индикаторы): ПК-2 (ПК-2.2).

## Экспертное заключение

Представленный фонд оценочных средств (далее – ФОС) по дисциплине «Машинное обучение и анализ данных» соответствует требованиям ФГОС ВО.

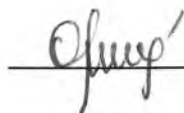
Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины представлены в полном объеме.

Виды оценочных средств, включенные в представленный фонд, отвечают основным принципам формирования ФОС.

Разработанный и представленный для экспертизы фонд оценочных средств рекомендуется к использованию в процессе подготовки обучающихся по указанному направлению 09.04.01 Информатика и вычислительная техника.

Председатель учебно-методической  
комиссии Краснодонского факультета  
инженерии и менеджмента (филиала).



Родионова О.Ю.

### Лист изменений и дополнений

| №<br>п/п | Виды дополнений и<br>изменений | Дата и номер протокола<br>заседания кафедры<br>(кафедр), на котором были<br>рассмотрены и одобрены<br>изменения и дополнения | Подпись<br>(с расшифровкой)<br>заведующего кафедрой<br>(заведующих кафедрами) |
|----------|--------------------------------|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
|          |                                |                                                                                                                              |                                                                               |
|          |                                |                                                                                                                              |                                                                               |
|          |                                |                                                                                                                              |                                                                               |
|          |                                |                                                                                                                              |                                                                               |