**МИНИСТЕРСТВО науки и высшего образования Российской Федерации**

федеральное Государственное бюджетное образовательное учреждение высшего образования

**«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ университет»**

**Утверждаю**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_\_\_» \_\_\_\_\_\_\_ 202\_ г.

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

**для проверки сформированности компетенций**

|  |  |
| --- | --- |
| дисциплины: | **Машинное обучение и анализ данных** |
| направление подготовки: | **01.04.02 Прикладная математика и информатика** |
| Направленность (профиль): | **Машинное обучение и анализ данных** |
| форма обучения: | **очная, очно-заочная** |

Фонд оценочных средств рассмотрен

на заседании кафедры математики и прикладных информационных технологий

**ПРОЦЕСС ИЗУЧЕНИЯ ДИСЦИПЛИНЫ НАПРАВЛЕН НА ФОРМИРОВАНИЕ   
СЛЕДУЮЩИХ КОМПЕТЕНЦИЙ**

|  |  |
| --- | --- |
| Код и наименование компетенции | Код и наименование индикатора достижения компетенции |
| 1 | 2 |
| ОПК-2. Способен совершенствовать и реализовывать новые математические методы решения прикладных задач | ОПК-2.1. Способен получать информацию о новых математических методах решения прикладных задач  ОПК-2.2. Способен совершенствовать и реализовывать новые математические методы для решения задач в профессиональной деятельности |
| ОПК-3. Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности | ОПК-3.1. Способен анализировать проблемы и тенденции разработки математических моделей для решения задач в профессиональной деятельности  ОПК-3.2. Способен разрабатывать математические модели для решения прикладных задач и их использования в профессиональной деятельности |

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕРКИ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ**

***Для оценки ОПК-2*** Способен совершенствовать и реализовывать новые математические методы решения прикладных задач

1) Какие из следующих утверждений верны в отношении деревьев решений? Выберите **несколько правильных** ответов.

1. Деревья решений могут обрабатывать как числовые, так и категориальные данные.
2. Деревья решений склонны к переобучению, если не применять методы регуляризации.
3. Деревья решений всегда находят оптимальное решение.
4. Деревья решений могут быть легко интерпретированы и визуализированы.
5. Деревья решений чувствительны к масштабированию признаков

**Правильный ответ: 1, 2, 4.**

2) Какие методы могут помочь предотвратить переобучение в деревьях решений? Выберите **несколько правильных** ответов.

1. Обрезка дерева (Pruning)

2. Ограничение глубины дерева (Max Depth)

3. Минимальное количество образцов в листе (Min Samples Leaf)

4. Использование ансамблей (например, Random Forest, Gradient Boosting)

5. Нормализация данных

**Правильный ответ: 1, 2, 3, 4.**

3) Укажите правильную последовательность шагов в алгоритме k-means?

1. Расчет центроидов

2. Присвоение точек кластерам

3. Инициализация центроидов

4. Проверка на сходимость

**Правильный ответ: 3, 2, 1, 4.**

4) Выберите **несколько правильных** ответов. Какие метрики расстояния обычно используются в kNN?

1. Евклидово расстояние

2. Расстояние Манхеттэна

3. Косинусное сходство

4. Расстояние Миньковского

**Правильный ответ: 1, 2, 3**

5) Какие из перечисленных методов относятся к ансамблевым методам обучения? Выберите **несколько правильных** ответов.

1. Линейная регрессия

2. Случайный лес (Random Forest)

3. Метод k-ближайших соседей (k-NN)

4. Градиентный бустинг (Gradient Boosting)

5. Наивный Байес

6. Экстремальные деревья (Extremely Randomized Trees)

**Правильный ответ: 2, 4, 6.**

6) Укажите правильную последовательность типичных шагов для построения ансамбля моделей методом бэггинга (bagging)?

1. Обучение базовых моделей на отдельных подвыборках данных.

2. Объединение предсказаний базовых моделей для получения итогового результата (например, усреднением или голосованием).

3. Генерация случайных подвыборок данных с возвращением (bootstrap samples).

4. Оценка качества ансамбля на тестовой выборке.

**Правильный ответ: 3, 1, 2, 4.**

7) Установите соответствие.

К каждой позиции, данной в левом столбце, подберите соответствующую позицию из правого столбца.

|  |  |  |  |
| --- | --- | --- | --- |
| Алгоритм | | Описание алгоритма машинного обучения | |
| А | Алгоритм линейной регрессии | 1 | Метод, который строит модель в виде графа, принимая решения на каждом узле |
| Б | Метод K-ближайших соседей (KNN) | 2 | Метод, который основан на теореме Байеса и предполагает независимость признаков. |
| В | Деревья решений | 3 | Метод, который предсказывает числовые значения на основе линейной зависимости |
| Г | Наивный байесовский классификатор | 4 | Алгоритм, который ищет гиперплоскость для разделения классов с максимальным зазором |
| Д | Метод опорных векторов (SVM) | 5 | Метод, который классифицирует объект на основе классов его ближайших объектов |

Запишите выбранные цифры под соответствующими буквами.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| А | Б | В | Г | Д |
|  |  |  |  |  |

**Правильный ответ:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **А** | **Б** | **В** | **Г** | **Д** |
| **3** | **5** | **1** | **2** | **4** |

8) Какие из перечисленных алгоритмов машинного обучения используются для решения задач классификации. Выберите **несколько правильных** ответов.

1. Линейная регрессия (Linear Regression)

2. Регрессия Лассо (Lasso Regression)

3. Метод к ближайших соседей (k-NN)

4. Деревья решений (Decision Tree)

5. Алгоритм Apriori (Apriori)

6. Метод опорных векторов (SVM)

7. Наивный байесовский классификатор Naïve Bayes)

8. Логистическая регрессия (Logistic Regression)

**Правильный ответ: 3, 4, 6, 7, 8**

9) Какой алгоритм помогает корректировать веса нейронов в ходе обучения?

**Правильный ответ: алгоритм обратного распространения ошибки**

10) Выберите **1 правильный** вариант ответа. Какой метод кросс-валидации наиболее устойчив к переобучению и обеспечивает наиболее надёжную оценку обобщающей способности модели?

1. K-кратная перекрестная проверка (K-fold cross-validation)

2. Простая перекрестная проверка (Leave-One-Out cross-validation)

3. Повторяющаяся k-кратная перекрестная проверка (Repeated k-fold cross-validation)

4. Перекрестная проверка с перемешиванием (Shuffle-split cross-validation)

**Правильный ответ: 3. Повторяющаяся k-кратная перекрестная проверка (Repeated k-fold cross-validation)**

***Для оценки ОПК-3*** Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности

1) Установите соответствие.

К каждой позиции, данной в левом столбце, подберите соответствующую позицию из правого столбца.

|  |  |  |  |
| --- | --- | --- | --- |
| Тип задачи машинного обучения | | Практическая задача | |
| А | Классификация | 1 | Имеются данные о ценах на недвижимость и их характеристиках (площадь, количество комнат, местоположение и т.д.). Нужно создать модель, которая будет предсказывать цену нового дома на основе его характеристик. |
| Б | Регрессия | 2 | Требуется сгруппировать клиентов компании в различные сегменты на основе их покупательского поведения. |
| В | Кластеризация | 3 | Необходимо обнаружить мошеннические транзакции в платежной системе. В системе имеются данные о всех транзакциях, и нужно идентифицировать транзакции, которые отличаются от обычных. |
| Г | Обнаружение аномалий | 4 | Требуется предсказать, купит ли клиент новый продукт вашей компании на основе его демографических данных и истории покупок. |

Запишите выбранные цифры под соответствующими буквами.

|  |  |  |  |
| --- | --- | --- | --- |
| А | Б | В | Г |
|  |  |  |  |

**Правильный ответ:**

|  |  |  |  |
| --- | --- | --- | --- |
| **А** | **Б** | **В** | **Г** |
| **4** | **1** | **2** | **3** |

2) Установите соответствие.

К каждой позиции, данной в левом столбце, подберите соответствующую позицию из правого столбца.

|  |  |  |  |
| --- | --- | --- | --- |
| Область знаний  в машинном обучении | | Входные данные | |
| А | Data Science | 1 | Данные и алгоритмы |
| Б | Data Mining | 2 | Признаки и алгоритмы |
| В | Classical Programming | 3 | Данные и признаки |

Запишите выбранные цифры под соответствующими буквами.

|  |  |  |
| --- | --- | --- |
| А | Б | В |
|  |  |  |

**Правильный ответ:**

|  |  |  |
| --- | --- | --- |
| **А** | **Б** | **В** |
| **3** | **1** | **2** |

3) Вы обучаете модель классификации изображений, которая должна отличать кошек от собак. Какой тип машинного обучения вы используете?

**Правильный ответ: обучение с учителем**

4) Выберите **1 правильный** вариант ответа. Вы тренируете модель для предсказания цен на недвижимость. Какой показатель качества модели наиболее важен в этом случае?

1. Точность (Accuracy)

2. Среднеквадратичная ошибка (MSE)

3. F1-мера

4. AUC-ROC

**Правильный ответ: 2. Среднеквадратичная ошибка (MSE)**

5) Выберите **1 правильный** вариант ответа. Что такое переобучение в машинном обучении?

1. Модель слишком хорошо обобщает данные и предсказывает новые данные с высокой точностью.

2. Модель не может найти закономерности в данных и предсказывает случайные значения.

3. Модель слишком хорошо запоминает обучающие данные и плохо обобщает на новые данные.

4. Модель не имеет достаточного количества данных для обучения.

**Правильный ответ: 3. Модель слишком хорошо запоминает обучающие данные и плохо обобщает на новые данные.**

6) Укажите правильную типичную последовательность шагов в процессе разработки моделей машинного обучения?

1. Выбор модели

2. Сбор и подготовка данных

3. Валидация модели

4. Обучение модели

5. Развертывание и мониторинг

6. Тестирование модели

**Правильный ответ: 2, 1, 4, 3, 6, 5**

7) Выберите **несколько правильных** ответов. **Какие из следующих алгоритмов можно использовать для уменьшения размерности и визуализации данных?**

1. KNN

2. Apriori

3. t-SNE

4. Наивный байесовский классификатор

5. PCA

**Правильный ответ: 3, 5**

8)Выберите **1 правильный** ответ.Какую проблему можно решить с помощью логистической регрессии?

1. Прогнозирование непрерывной переменной, такой как температура.

2. Классификация объектов на две или более категорий.

3. Выявление скрытых закономерностей в больших наборах данных.

4. Прогнозирование временных рядов.

**Правильный ответ: 2. Классификация объектов на две или более категорий.**

9) Какая метрика наиболее подходит для оценки качества классификатора, если классы сильно не сбалансированы?

**Правильный ответ: F1-мера**

10)Вы тренируете модель регрессии и замечаете высокую дисперсию. Какой из методов может помочь уменьшить дисперсию?

**Правильный ответ: регуляризация**

**Критерии оценивания уровня сформированности компетенций**

Оценка «отлично» выставляется обучающемуся, если им даны правильные ответы более, чем на 90% вопросов.

Оценка «хорошо» выставляется обучающемуся, если им даны правильные ответы на 760-90% вопросов.

Оценка «удовлетворительно» выставляется обучающемуся, если им даны правильные ответы на 61-75% вопросов.

Оценка «неудовлетворительно» выставляется обучающемуся, если им даны правильные ответы менее чем на 61% вопросов.