

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ
ТЕХНОЛОГИИ СБОРА И ПОДГОТОВКИ ДАННЫХ В АГРОИНЖЕНЕРИИ

Направление подготовки: 35.03.06 Агроинженерия

Профиль подготовки: Искусственный интеллект

Квалификация (степень)
выпускника: Бакалавр

Череповец, Вологда – Молочное
2023

1 Паспорт фонда оценочных средств по дисциплине (модулю)

1.1 Текущий контроль

№ п/п	Раздел (тема) дисциплины	Результаты обучения (компетенции)	Наименование оценочного средства / Форма текущего контроля	Метод контроля
1.	Технологии сбора и подготовки данных в агроинженерии	ПК-11, ПК-17, ПК-18	Лабораторное занятие	Устный опрос
2.	Трансформация и визуализация данных	ПК-11, ПК-17, ПК-18	Лабораторное занятие	Устный опрос
3.	Очистка и предобработка данных	ПК-11, ПК-17, ПК-18	Лабораторное занятие	Устный опрос

1.2 Промежуточная аттестация

В соответствии с учебным планом промежуточная аттестация по учебной дисциплине «Машинное обучение в агроинженерии» предусматривает проведение зачета. Для оценки результатов обучения используется метод тестирования.

**2 Комплект оценочных материалов для проведения текущего контроля
оценки знаний, умений и уровня сформированности компетенций**

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

**Вопросы для контроля освоения компетенции
УК-11**

«Способен планировать и организовывать свою деятельность в цифровом пространстве с учетом правовых и этических норм взаимодействия человека и искусственного интеллекта и требований информационной безопасности»

1. Почему необходима предварительная обработка данных
2. Что включает в себя очистка данных?
3. Какие методы можно применить для восстановления численных значений
4. Как можно восстановить категориальные значения?
5. Что такое внутренне обогащение данных
6. Когда необходимо внешнее обогащение данных?
7. Что можно сказать о данных, имеющих совпадающие значения центральных средних?
8. В каком случае удобнее использовать медиану, чем среднее арифметическое.
9. Техники визуализации данных.
10. Визуализация данных: подтверждающий анализ и представление результатов.
11. Какие тренды информационно-коммуникационных технологий способствовали развитию Data Mining!
12. Приведите примеры применения методов Data Mining для решения практических задач.
13. Какие области человеческой деятельности наиболее и наименее подходят для их анализа методами Data Mining?
14. Что понимается под Data Mining и Big Data? Почему возникла такая терминология?
15. Приведите примеры коммерческих многофункциональных систем и свободно распространяемых решений, реализующих инструментарий Data Mining. Их сравнительные характеристики.

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

**Вопросы для контроля освоения компетенции
ПК-17**

«Способен осуществлять сбор и подготовку данных для систем
искусственного интеллекта»

1. Проиллюстрируйте, что такое размах и межквартильный размах.
2. Поясните, что такое нулевая и альтернативная гипотезы.
3. На каком этапе используется поисковый анализ?
4. В чем различие визуальной информации и визуализации данных?
5. Какие диаграммы можно использовать в визуализациях для мониторинга?
6. На чем может основываться научная визуализация?
7. Приведите примеры диаграмм в геометрической технике.
8. Какие диаграммы подходят для визуализации категориальных данных?
9. Какие диаграммы подойдут для визуализации числовых данных?
10. Связи между какими признаками можно анализировать с помощью таблиц сопряженности?
11. На основании какого коэффициента можно говорить о возможном наличии линейных связей между количественными признаками.?
12. Каким образом формулируются нулевая и альтернативная гипотезы?
13. В чем заключается оптимизация признакового пространства?
14. Какие методы с трансформацией и без трансформации пространства применяют и в чем их отличия?
15. Что такое визуализация и какие инструменты ее реализации существуют?
16. Что такое обнаружение аномалий? Приведите примеры применения этого подхода и методы его реализации.

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

Вопросы для контроля освоения компетенции
ПК-18
«Способен выполнять анализ больших данных»

1. Назовите возможные источники больших данных
2. Что понимается под интеллектуальным анализом данных?
3. Что включает в себя наука о данных (Data Science)
4. Какие стандарты данных вам известны?
5. Перечислите этапы реализации стандарта CRISP DM.
6. С помощью каких тестов проверяются гипотезы о среднем арифметическом?
7. С помощью каких тестов проверяются гипотезы о связи между качественными признаками?
8. На каком этапе анализа данных необходимо исследование статистических связей.
9. Приведите пример подтверждающего анализа.
10. Что такое статистические связи?
11. Какие классификации методов Data Mining существуют? Приведите примеры.
12. Каковы основные этапы интеллектуального анализа данных?
13. В чем состоит суть индуктивных и дедуктивных подходов в Data Mining?
14. В чем заключается предварительная обработка данных и какова ее цель? Какие подходы при этом применяются?
15. Какие инструменты, модели и технологии существуют сегодня для реализации высокопроизводительных вычислений? Какие критерии эффективности при этом используют?

3 Комплект оценочных материалов для проведения промежуточной аттестации по итогам изучения учебной дисциплины (модуля)

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

**Тесты для контроля освоения компетенции
УК-11**

«Способен планировать и организовывать свою деятельность в цифровом пространстве с учетом правовых и этических норм взаимодействия человека и искусственного интеллекта и требований информационной безопасности»

Какая из перечисленных программных систем не может быть использована для анализа данных?

Matlab

EXCEL

3dMax

Statistica

Специалист в области анализа и моделирования – это ...

эксперт

аналитик

инженер

Обнаружение в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности – это ...

распознавание образов

классификация

Data Mining

кластеризация

Объем вычислений, который при реализации алгоритма должен обеспечивать его выполнение за разумное время и по возможности быть нетребовательным к ресурсам компьютера

производительность

устойчивость

точность

Определение методами интеллектуального анализа значения непрерывного числового параметра на основании значений независимых переменных производится при решении задачи:

классификации

кластеризации

регрессии

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

**Тесты для контроля освоения компетенции
ПК-17**

«Способен осуществлять сбор и подготовку данных для систем
искусственного интеллекта»

Объективными предпосылками появления ошибки модели регрессии являются (выбрать два варианта):

нерепрезентативность выборки
вероятность ошибочной интерпретации переменных
вероятность ошибочного измерения переменных
сложность модели

Объект или описание объекта, системы для замещения (при определенных условиях, предположениях, гипотезах) одной системы (то есть оригинала) другой системой для лучшего изучения оригинала или воспроизведения каких-либо его свойств:

модель
макет
образ

Модели, полученные на основе эмпирических фактов, зависимостей называются ...

эмпирические
теоретические
смешанные

Модели, полученные на основе математических описаний, законов называются ...

эмпирические
теоретические
смешанные

Модель должна иметь хотя бы один параметр, изменениями которого можно имитировать поведение моделируемой системы в различных условиях. Это свойство называется ...

управляемость
замкнутость
адекватность
целостность
приближенность
упрощенность

Данные, определенным образом упорядоченные и организованные с целью обеспечения возможности применения к ним некоторых действий (например, визуального или машинного анализа) называют ...

неструктурированными

случайными

детерминированными

структурированными

Разбиение множества объектов или наблюдений на заданные группы (классы), внутри каждой из которых они предполагаются похожими друг на друга, имеющими примерно одинаковые свойства и признаки

классификация

регрессия

кластеризация

ассоциация

Установление зависимости непрерывной выходной переменной от входных переменных

классификация

регрессия

кластеризация

ассоциация

Объединение объектов или наблюдений в непересекающиеся группы, на основе близости значений их признаков

классификация

регрессия

кластеризация

ассоциация

Выявление закономерностей между связанными событиями

классификация

регрессия

кластеризация

ассоциация

Набор данных, каждая запись которого представляет собой обучающий пример, содержащий заданное входное воздействие, и соответствующий ему правильный выходной результат

обучающая выборка

представительная выборка

тестовая выборка

генеральная совокупность

Ошибка, допущенная моделью на обучающем множестве

ошибка обучения

ошибка измерения

ошибка обобщения

Ошибка, полученная на тестовых примерах, то есть вычисляемая по тем же формулам, но для тестового множества

ошибка обучения
ошибка измерения
ошибка обобщения

Комплекс методов и алгоритмов, применяемых в аналитическом приложении с целью подготовки данных к решению конкретной задачи и приведения их в соответствие с требованиями, определяемыми спецификой задачи и способами ее решения

предобработка данных
очистка данных
консолидация данных
трансформация данных
визуализация данных

Мера статистической связи между функцией и ее копией, сдвинутой на некоторый интервал

автокорреляция
дисперсия
среднее значение

Массивы данных, объемы которых настолько велики, что их обработка традиционными средствами становится неэффективной или невозможной

большие данные
структурированные данные
случайные данные

Предварительный анализ данных с целью выявления наиболее общих зависимостей, закономерностей и тенденций, предназначенный для разработки наилучшей стратегии углубленного анализа и выдвижения гипотез

регрессионный анализ
разведочный анализ
дисперсионный анализ
корреляционный анализ

Набор статистических процедур для изучения зависимостей между случайными переменными. В аналитических технологиях Data Mining широко используется для решения задач прогнозирования, оценивания, классификации

регрессионный анализ
разведочный анализ
дисперсионный анализ
корреляционный анализ

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

**Тесты для контроля освоения компетенции
ПК-18**

«Способен выполнять анализ больших данных»

Укажите правильную последовательность этапов Knowledge Discovery in Databases – процесса обнаружения знаний в базах данных:
трансформация, интерпретация результатов, выборка, очистка, построение моделей.

построение моделей, выборка, очистка, трансформация, интерпретация результатов

построение моделей, выборка, очистка, трансформация, интерпретация результатов

выборка, очистка, трансформация, построение моделей, интерпретация результатов

Какая из перечисленных задач не относится к задачам Data Mining:

классификация

кластеризация

решение дифференциальных уравнений

прогнозирование

Основное назначение систем интеллектуального анализа (Data Mining):

обнаружение в сырых данных скрытых знаний

проведение статистического анализа

решения задач математического программирования

поиск агрегированных данных

С чем может быть связано получение неверного результата при программной реализации вычислительного метода анализа данных?

Укажите неправильный ответ

неверно выбран язык программирования

ошибка в программном коде

неверно выбраны параметры метода

неправильно построен алгоритм решения задачи

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное
образовательное учреждение высшего образования
«Череповецкий государственный университет»
Институт информационных технологий

Кафедра «Математического и программного обеспечения ЭВМ»

**Вопросы к зачету для контроля освоения компетенций ПК-11, ПК-17,
ПК-18**

1. Применение БПЛА для сбора данных.
2. Коммуникационные технологии Agro IoT.
3. Хранилища данных.
4. Принципы анализа данных.
5. Структурированные и неструктурированные данные.
6. Подготовка данных к анализу.
7. Применение технологии KDD для анализа агротехнических данных.
8. Применение технологии Data Mining для анализа агротехнических данных.
9. Инструменты сбора и анализа данных: POWER BI.
10. Инструменты сбора и анализа данных: MS EXCEL.
11. Инструменты сбора и анализа данных: ORANGE DATA MINING, Python и его библиотеки.
12. Инструменты сбора и анализа данных: Python и его библиотеки.
13. Понятие трансформации данных.
14. Трансформация упорядоченных данных.
15. Группировка данных.
16. Слияние данных.
17. Квантование.
18. Нормализация и кодирование данных.
19. Введение в визуализацию.
20. Визуализаторы общего назначения.
21. OLAP-анализ.
22. Визуализаторы для анализа и оценки качества моделей.
23. Визуализаторы, применяемые для интерпретации результатов анализа.
24. Оценка качества данных.
25. Технологии и методы оценки качества данных.
26. Очистка и предобработка.
27. Фильтрация данных.
28. Обработка дубликатов и противоречий.
29. Выявление аномальных значений.
30. Восстановление пропущенных значений.
31. Введение в сокращение размерности.

32. Сокращение числа признаков.
33. Сокращение числа значений признаков и записей
34. Сэмплинг.
35. Структурированные и неструктурированные данные.
36. Анализ больших данных.

Фонд оценочных средств составлен в соответствии с требованиями ФГОС ВО и с учетом рекомендаций ОПОП ВО по направлению подготовки 35.03.06 Агроинженерия.

Разработчик: канд. техн. наук, доцент Шабалов В.А

Фонд оценочных средств одобрен на заседании кафедры математического и программного обеспечения ЭВМ 25 сентября 2023 года, протокол № 2.

Зав. кафедрой: Зав. кафедрой: доктор техн. наук, профессор Ершов Е.В.