

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Гаранин Максим Алексеевич

Должность: Ректор

Дата подписания: 14.03.2024 09:29:11

Уникальный программный ключ:

7708e5a47e65a8ee02711b298d7c78bd1e40b936

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ

Приложение
к рабочей программе дисциплины

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

Работа с большими данными и машинное обучение *(наименование дисциплины)*

Направление

09.04.02 Информационные системы и технологии
(код и наименование)

Направленность (профиль)

Корпоративные информационные системы
(наименование)

Содержание

1. Пояснительная записка.
2. Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций.
3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации.

1. Пояснительная записка

Цель промежуточной аттестации – оценивание промежуточных и окончательных результатов обучения по дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы.

Формы промежуточной аттестации: *зачет, семестр 3.*

Перечень компетенций, формируемых в процессе освоения дисциплины

Код и наименование компетенции	Код индикатора достижения компетенции
УК-2: Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2
ОПК-2: Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.2
ОПК-7: Способен разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений	ОПК-7.2

Результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы (семестр 4)
УК-2.2: Анализирует этапы жизненного цикла проекта, этапы разработки и реализации проекта, методы разработки и управления проектами	Обучающийся знает: особенности этапов жизненного цикла проекта, разработки и реализации проекта, методы разработки и управления проектами	Вопросы (№1 - №7)
	Обучающийся умеет: анализировать этапы жизненного цикла проекта, этапы разработки и реализации проекта, методы разработки и управления проектами	Задание (№1 - №4)
	Обучающийся владеет: навыками анализа этапов жизненного цикла проекта, разработки и реализации проекта, методы разработки и управления проектами	

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы (семестр 4)
ОПК-2.2: Использует интеллектуальные технологии для решения профессиональных задач	Обучающийся знает: особенности, принципы использования интеллектуальных технологий при решении задач, связанных с большими данными	Вопросы (№8 - №14)
	Обучающийся умеет: использовать интеллектуальные технологии для решения профессиональных задач	Задание (№5 - №7)
	Обучающийся владеет: навыками использования интеллектуальных технологий для решения профессиональных задач	

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы (семестр 4)
ОПК-7.2: Строит математические модели для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений	Обучающийся знает: особенности построения математических моделей для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений на основе больших данных	Вопросы (№15 - №20)
	Обучающийся умеет: строить математические модели для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений на основе больших данных	Задание (№8 - №11)
	Обучающийся владеет: навыками построения математических моделей для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений на основе больших данных	

Промежуточная аттестация (зачет) проводится в одной из следующих форм:

- 1) собеседование;
- 2) выполнение заданий в ЭИОС СамГУПС.

2. Типовые¹ контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций

2.1 Типовые вопросы (тестовые задания) для оценки знаниевого образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
УК-2.2: Анализирует этапы жизненного цикла проекта, этапы разработки и реализации проекта, методы разработки и управления проектами	Обучающийся знает: особенности этапов жизненного цикла проекта, разработки и реализации проекта, методы разработки и управления проектами
ОПК-2.2: Использует интеллектуальные технологии для решения профессиональных задач	Обучающийся знает: особенности, принципы использования интеллектуальных технологий при решении задач, связанных с большими данными
ОПК-7.2: Строит математические модели для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений	Обучающийся знает: особенности построения математических моделей для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений на основе больших данных
<p><i>Примеры вопросов</i></p> <p>1. Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк PyTorch TensorFlow Scikit-learn Flask</p>	

¹ Приводятся типовые вопросы и задания. Оценочные средства, предназначенные для проведения аттестационного мероприятия, хранятся на кафедре в достаточном для проведения оценочных процедур количестве вариантов. Оценочные средства подлежат актуализации с учетом развития науки, образования, культуры, экономики, техники, технологий и социальной сферы. Ответственность за нераспространение содержания оценочных средств среди обучающихся университета несут заведующий кафедрой и преподаватель – разработчик оценочных средств.

2. Формат Parquet считается
колоночным (столбцовым)

строковым

неструктурированным

полуструктурированным

3. Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью

Apache Hive

Apache AirFlow

Apache Kafka

Apache Hadoop

4. Выберите технологию потоковой обработки событий в режиме реального времени

Apache Kafka

Spark Streaming

Apache Hadoop

MapReduce

5. Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов

Нельзя

Можно

6. Повысить производительность Apache Kafka можно с помощью:

Увеличения размера сообщений

Замены HDD-дисков на SSD

Увеличения плотности разделов на каждом брокере

Повышения коэффициента репликации

7. Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит

Apache Hadoop

Apache Kafka

Apache AirFlow

Apache Spark

8. Apache NiFi используется для

оптимизации SQL-запросов к DWH

маршрутизации потоков Big Data и построения ETL-конвейеров

визуализации результатов аналитики

эффективного хранения больших данных

9. Для машинного обучения подходят данные

Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов

Любых форматов в цифровом виде

Бинарные

Числовые типа int

10. Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД

Elasticsearch

Hive

Cassandra

HBase

11. Что, из ниже перечисленного, относится к обучающей выборке?

классификация данных

объекты с известными ответами

алгоритм решающий функцию

12. Объекты состоят из признаков?

Да

Нет

13. Что называют данными в машинном обучении?

матрицы

объекты

признаки

<p>алгоритм</p> <p>функция</p> <p>14 Выберите правильный ответ. Задача классификации - это:</p> <p>множество объектов, разделенных на классы</p> <p>исследование влияние одного или нескольких признаков на объект</p> <p>определение порядка признака согласно рангу</p> <p>15 Выберите правильный ответ. Задача регрессии - это:</p> <p>множество объектов, разделенных на классы</p> <p>исследование влияние одного или нескольких признаков на объект</p> <p>определение порядка признака согласно рангу</p> <p>16 Кроме рисования искусственный интеллект научился разбираться в музыке. Но насколько хорошо работает программа по определению музыкальных стилей?</p> <p>Да, лучше, чем программа, написанная вручную</p> <p>Да, но программа написанная вручную будет точнее</p> <p>Нет, в режиме реального времени программа не справится</p> <p>17 Способен ли ИИ сравниться с человеческим интеллектом в общей совокупности способностей?</p> <p>Нет, пока этот уровень недостижим</p> <p>Нет, но технология уже близка к уровню мозга человека</p> <p>Да, и уже идёт работа над созданием искусственного сверхума</p> <p>18 К какому времени относятся первые упоминания об искусственно созданных человекоподобных существах?</p> <p>Период до н.э. (в источниках Древнего Египта)</p> <p>Период Возрождения (в трудах Да Винчи)</p> <p>XX век (в материалах, появившихся после возникновения обучающих алгоритмов)</p> <p>19 Новая программа от Google научилась рисовать на основе эскизов, сделанных людьми. Что при этом учитывала программа?</p> <p>Стиль и цветовую гамму, типичную для похожих изображений</p> <p>Только конечный результат</p> <p>Концепцию (идею) рисунка</p> <p>20 Выберите правильный ответ. Задача ранжирования - это:</p> <p>множество объектов, разделенных на классы</p> <p>исследование влияние одного или нескольких признаков на объект</p> <p>определение порядка признака согласно рангу</p>
--

2.2 Типовые задания для оценки навыкового образовательного результата

Проверяемый образовательный результат:

<u>Код и наименование индикатора достижения компетенции</u>	<u>Образовательный результат</u>
УК-2.2: Анализирует этапы жизненного цикла проекта, этапы разработки и реализации проекта, методы разработки и управления проектами	Обучающийся умеет: анализировать этапы жизненного цикла проекта, этапы разработки и реализации проекта, методы разработки и управления проектами
	Обучающийся владеет: навыками анализа этапов жизненного цикла проекта, разработки и реализации проекта, методы разработки и управления проектами
ОПК-2.2: Использует интеллектуальные технологии для решения профессиональных задач	Обучающийся умеет: использовать интеллектуальные технологии для решения профессиональных задач
	Обучающийся владеет: навыками использования интеллектуальных технологий для решения профессиональных задач
ОПК-7.2: Строит математические модели для реализации успешного функционирования распределенных	Обучающийся умеет: строить математические модели для реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений на основе больших данных
	Обучающийся владеет: навыками построения математических моделей для

информационных систем и систем поддержки принятия решений	реализации успешного функционирования распределенных информационных систем и систем поддержки принятия решений на основе больших данных
<p><i>Примеры заданий Реализовать в любой среде программирования следующую задачу:</i></p> <ol style="list-style-type: none"> 1. Линейная модель классификации. Отступ. Обучение линейных классификаторов через верхнюю оценку на долю ошибок. Примеры верхних оценок. 2. Логистическая регрессия. Оценивание вероятностей. 3. Метод опорных векторов. Вывод постановки задачи для разделимого и неразделимого случаев. Использование ядер в методе опорных векторов. Теорема Мерсера. 4. Многоклассовая классификация: one-vs-all, all-vs-all. Многоклассовая логистическая регрессия. Микро и макро – усреднения. 5. Решающие деревья: определение и жадный алгоритм обучения. Функционал качества при выборе предиката. Общий вид критерия информативности (через функцию потерь) и конкретные примеры для регрессии (дисперсия) и классификации (критерий Джини и энтропийный критерий). 6. Композиции алгоритмов. Разложение ошибки на смещение и разброс (с выводом). 7. Бэггинг и случайные леса. 8. Градиентный бустинг. Обучение базовых алгоритмов для произвольной дифференцируемой функции потерь. Сокращение шага. 9. Нейронные сети, их обучение методом обратного распространения ошибки. 10. Задача кластеризации. Карты Кахоннена. 11. Нелинейные методы классификации: наивный байесовский классификатор, метод Kmeans 	

2.3. Перечень вопросов для подготовки обучающихся к промежуточной аттестации

12. Основные определения в машинном обучении: объект, целевая переменная, признак, модель, обучающая выборка, функционал ошибки, обучение, переобучение. Приведите примеры задач обучения с учителем и без учителя.
13. Линейная модель регрессии. Аналитическое решение для среднеквадратичной ошибки (с выводом). Градиентное обучение линейной регрессии.
14. Функционалы ошибки для регрессии: MSE, MAE, коэффициент детерминации. MSE как максимум правдоподобия.
15. Градиентные методы обучения. Свойство градиента о направлении наискорейшего убывания. Градиентный спуск. Методы оценивания градиента.
16. Сингулярное разложение матриц. Сингулярные числа. Сингулярное разложение в задаче регрессии со среднеквадратичной ошибкой.
17. регуляризация. Аналитический вид вектора весов в линейной регрессии со среднеквадратичной ошибкой и L-2 регуляризатором (с выводом).
18. L1-регуляризация. Почему использование L1-регуляризатора приводит к отбору признаков?
19. Линейная модель классификации. Отступ. Обучение линейных классификаторов через верхнюю оценку на долю ошибок. Примеры верхних оценок.
20. Функционалы ошибки для классификации: матрица ошибок, accuracy, precision, recall, F-мера. ROC-кривая и AUC-ROC. Precision-recall-кривая и площадь под ней
21. Логистическая регрессия. Оценивание вероятностей.
22. Метод опорных векторов. Вывод постановки задачи для разделимого и неразделимого случаев.
23. Использование ядер в методе опорных векторов. Теорема Мерсера.
24. Многоклассовая классификация: one-vs-all, all-vs-all. Многоклассовая логистическая регрессия. Микро и макро – усреднения.
25. Решающие деревья: определение и жадный алгоритм обучения. Функционал качества при выборе предиката. Общий вид критерия информативности (через функцию потерь) и конкретные примеры для регрессии (дисперсия) и классификации (критерий Джини и энтропийный критерий).
26. Композиции алгоритмов. Разложение ошибки на смещение и разброс (с выводом).
27. Бэггинг и случайные леса.
28. Градиентный бустинг. Обучение базовых алгоритмов для произвольной дифференцируемой функции потерь. Сокращение шага.
29. Нейронные сети, их обучение методом обратного распространения ошибки.
30. Задача кластеризации. Карты Кахоннена.
31. Нелинейные методы классификации: наивный байесовский классификатор, метод Kmeans

3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации

Критерии формирования оценок по ответам на вопросы, выполнению тестовых заданий

- оценка **«отлично»** выставляется обучающемуся, если количество правильных ответов на вопросы составляет 100 – 90% от общего объема заданных вопросов;
- оценка **«хорошо»** выставляется обучающемуся, если количество правильных ответов на вопросы – 89 – 76% от общего объема заданных вопросов;
- оценка **«удовлетворительно»** выставляется обучающемуся, если количество правильных ответов на тестовые вопросы – 75–60 % от общего объема заданных вопросов;
- оценка **«неудовлетворительно»** выставляется обучающемуся, если количество правильных ответов – менее 60% от общего объема заданных вопросов.

Критерии формирования оценок по результатам выполнения заданий

«Отлично/зачтено» – ставится за работу, выполненную полностью без ошибок и недочетов.

«Хорошо/зачтено» – ставится за работу, выполненную полностью, но при наличии в ней не более одной негрубой ошибки и одного недочета, не более трех недочетов.

«Удовлетворительно/зачтено» – ставится за работу, если обучающийся правильно выполнил не менее 2/3 всей работы или допустил не более одной грубой ошибки и двух недочетов, не более одной грубой и одной негрубой ошибки, не более трех негрубых ошибок, одной негрубой ошибки и двух недочетов.

«Неудовлетворительно/не зачтено» – ставится за работу, если число ошибок и недочетов превысило норму для оценки «удовлетворительно» или правильно выполнено менее 2/3 всей работы.

Виды ошибок:

- *грубые ошибки: незнание основных понятий, правил, норм; незнание приемов решения задач; ошибки, показывающие неправильное понимание условия предложенного задания.*
- *негрубые ошибки: неточности формулировок, определений; нерациональный выбор хода решения.*
- *недочеты: нерациональные приемы выполнения задания; отдельные погрешности в формулировке выводов; небрежное выполнение задания.*

Критерии формирования оценок по зачету

«Зачтено» - обучающийся демонстрирует знание основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем; приобрел необходимые умения и навыки, освоил вопросы практического применения полученных знаний, не допустил фактических ошибок при ответе, достаточно последовательно и логично излагает теоретический материал, допуская лишь незначительные нарушения последовательности изложения и некоторые неточности.

«Не зачтено» - выставляется в том случае, когда обучающийся демонстрирует фрагментарные знания основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем. У экзаменуемого слабо выражена способность к самостоятельному аналитическому мышлению, имеются затруднения в изложении материала, отсутствуют необходимые умения и навыки, допущены грубые ошибки и незнание терминологии, отказ отвечать на дополнительные вопросы, знание которых необходимо для получения положительной оценки.

