

**Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

Тульский филиал Финуниверситета

Кафедра «Математика и информатика»

СОГЛАСОВАНО

ЗАО «ЛИМ»

Генеральный директор

В.В. Куликов



«19» декабря 2024 г.

УТВЕРЖДАЮ

Директор филиала

Г.В. Кузнецов



«24» декабря 2024 г.

Е.В. Манохин

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ПО ДИСЦИПЛИНЕ
ПРИКЛАДНЫЕ ЗАДАЧИ МАШИННОГО ОБУЧЕНИЯ**

для студентов, обучающихся по направлению подготовки
09.03.02 Информационные системы и технологии,
образовательная программа «Информационные технологии в экономике»,
профиль «Информационные технологии в управлении цифровой
экономикой»

Рекомендовано Ученым советом Тульского филиала Финуниверситета
(протокол от 24 декабря 2024 г. № 22)

Одобрено заседанием кафедры «Математика и информатика»
(протокол от 02 декабря 2024 г. № 5)

Тула 2024

«Прикладные задачи машинного обучения»

КОД И НАИМЕНОВАНИЕ ОПОП:

09.03.02 Информационные системы и технологии

Образовательная программа - Информационные технологии в экономике

Профиль «Информационные технологии в управлении цифровой экономикой»

Общее количество заданий по дисциплине

Код компетенции	Наименование компетенции	Количество заданий
ПКП-3	Способен проводить исследования на всех этапах жизненного цикла программных систем	40
Всего:		40

Количество заданий по дисциплине в одном варианте

Код компетенции	Наименование компетенции	Количество заданий
ПКП-3	Способен проводить исследования на всех этапах жизненного цикла программных систем	20
Всего:		20

Типы заданий

Наименование типа заданий	Номер типа заданий
Задание комбинированного типа с выбором одного верного ответа из предложенных	1
Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных	2
Задание открытого типа	3
Задание открытого типа с развернутым ответом	4

Распределение заданий по компетенциям

Код компетенции	Наименование компетенции	Наименование индикаторов сформированности компетенции	Семестр	Номер задания
ПКП-3	Способен проводить исследования на всех этапах жизненного цикла программных систем	1. Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	7	1.1, 1.2, 2.1, 2.2, 3.1, 3.2, 4.1, 4.2, 5.1, 5.2, 6.1, 6.2, 7.1, 7.2, 8.1, 8.2, 9.1, 9.2, 10.1, 10.2
		2. Организует сбор и подготовку данных для систем машинного обучения.	7	11.1, 11.2, 12.1, 12.2, 13.1, 13.2, 14.1, 14.2, 15.1, 15.2, 16.1, 16.2, 17.1, 17.2, 18.1, 18.2, 19.1, 19.2, 20.1, 20.2

Типы, уровень сложности и время выполнения заданий

Код компетенции	Индикатор сформированности компетенции	Номер задания	Тип задания	Уровень сложности задания	Время выполнения задания (мин.)
ПКП-3	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	1.1, 1.2.	1	базовый	2
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	2.1, 2.2	1	базовый	2
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	3.1, 3.2	1	базовый	2
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	4.1, 4.2	1	базовый	2
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	5.1, 5.2	1	базовый	2
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	6.1, 6.2.	3	повышенный	4
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	7.1, 7.2	3	повышенный	4
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	8.1, 8.2	3	повышенный	4
	Демонстрирует знания основных методов машинного обучения и интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.	9.1, 9.2	4	высокий	8
	Демонстрирует знания основных методов машинного обучения и	10.1, 10.2	4	высокий	8

	интеллектуального анализа данных, применяет готовые инструменты для создания интеллектуальных алгоритмов.				
	Организует сбор и подготовку данных для систем машинного обучения.	11.1, 11.2.	1	базовый	2
	Организует сбор и подготовку данных для систем машинного обучения.	12.1, 12.2	1	базовый	2
	Организует сбор и подготовку данных для систем машинного обучения.	13.1, 13.2	2	базовый	2
	Организует сбор и подготовку данных для систем машинного обучения.	14.1, 14.2	1	базовый	2
	Организует сбор и подготовку данных для систем машинного обучения.	15.1, 15.2	1	базовый	2
	Организует сбор и подготовку данных для систем машинного обучения.	16.1, 16.2.	3	повышенный	4
	Организует сбор и подготовку данных для систем машинного обучения.	17.1, 17.2	3	повышенный	4
	Организует сбор и подготовку данных для систем машинного обучения.	18.1, 18.2	4	высокий	8
	Организует сбор и подготовку данных для систем машинного обучения.	19.1, 19.2	4	высокий	8
	Организует сбор и подготовку данных для систем машинного обучения.	20.1, 20.2	4	высокий	8

Задания, позволяющие осуществлять оценку компетенций, установленных для дисциплины

Вариант 1

Задание 1.1.

Какой метод машинного обучения используется для прогнозирования будущих значений на основе исторических данных?

- А) Кластеризация
- Б) Регрессия
- В) Метод опорных векторов
- Г) Дерево решений

Ответ:

Задание 2.1.

Какие данные используются для обучения моделей машинного обучения?

- А) Обучающие выборки.
- Б) Тестовые выборки.
- В) Контрольные выборки.
- Г) Любые доступные данные.

Ответ:

Задание 3.1.

Какой тип признаков необходимо кодировать перед обучением модели в машинном обучении?

- А. Числовые признаки
- Б. Категориальные признаки
- В. Признаки, содержащие пропуски
- Г. Логические признаки

Ответ:

Задание 4.1.

Какие задачи машинного обучения используются для обнаружения аномалий?

- А) Выявление мошенничества в финансовых транзакциях.
- Б) Анализ данных о продажах для определения тенденций.
- В) Создание рекомендательных систем.
- Г) Автоматическое распознавание речи.

Ответ:

Задание 5.1.

Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?

- А) Обучение с учителем
- Б) Обучение без учителя
- В) Глубинное обучение
- Г) Обучение с подкреплением

Ответ:

Задание 6.1.

Подраздел искусственного интеллекта, который позволяет системам самостоятельно обучаться на данных и делать предсказания или принимать решения без явного программирования на выполнение конкретной задачи – это ...

Ответ:

Задание 7.1.

Что такое ансамблевые методы в машинном обучении?

Ответ:

Задание 8.1.

Что называют данными в машинном обучении?

Ответ:

Задание 9.1.

Что такое машинное переобучение и как его избежать?

Ответ:

Задание 10.1.

Перечислите и охарактеризуйте не менее трех факторов, которые необходимо учитывать при выборе модели машинного обучения:

Ответ:

Задание 11.1.

Что, из ниже перечисленного, относится к обучающей выборке?

- А) классификация данных
- Б) объекты с известными ответами
- В) алгоритм, решающий функцию

Ответ:

Задание 12.1.

Дробление крупных кластеров на более мелкие называется задачей

- А) наименьших квадратов
- Б) сопряженного градиента
- В) аппроксимации
- Г) таксономии

Ответ:

Задание 13.1.

Какие задачи из ниже перечисленных относятся к задачам классификации?

- А) определение наиболее целесообразного способа лечения;
- Б) определение длительности и исхода заболевания;
- В) оценивание кредитоспособности заёмщика
- Г) задачи поискового вывода

Ответ:

Задание 14.1.

Какой тип экспериментального исследования имеет цель - либо решение конкретной прикладной задачи, либо выявление «слабых мест»?

- А) исследование задач ранжирования
- Б) исследование на реальных данных
- В) исследование на модельных данных

Ответ:

Задание 15.1.

По контрольной выборке вычисляется:

- А) контрольная выборка
- Б) эмпирическая оценка среднего риска
- В) тестируемый алгоритм
- Г) уровень сложности алгоритма

Ответ:

Задание 16.1.

Приведите не менее трех примеров использования машинного обучения в прикладных задачах.

Ответ:

Задание 17.1.

Как называется процесс подачи данных в модель машинного обучения для корректировки ее внутренних параметров и повышения производительности?

Ответ:

Задание 18.1.

Как работают рекомендательные системы на основе машинного обучения?

Ответ:

Задание 19.1.

Укажите и охарактеризуйте не менее трех основных типов задач машинного обучения?

Ответ:

Задание 20.1.

В чем заключается разница между тестовым и валидационным наборами данных?

Ответ:

Вариант 2

Задание 1.2.

Какая задача машинного обучения заключается в определении принадлежности объекта к одному из заранее определённых классов?

- А) Регрессия
- Б) Прогнозирование
- В) Классификация
- Г) Кластеризация

Ответ:

Задание 2.2.

Какой метод используется для уменьшения размерности данных?

- А) Регрессия
- Б) К-ближайших соседей
- В) РСА (главные компоненты)
- Г) Метод опорных векторов

Ответ:

Задание 3.2.

Что такое «обучение с учителем» в контексте машинного обучения?

- А) Автоматизированное тестирование программ
- Б) Создание искусственных нейронных сетей
- В) Обучение модели на основе размеченных данных с известными выходами

Г) Анализ временных рядов

Ответ:

Задание 4.2.

Что представляет собой метод «решающие деревья»?

- А) Алгоритм формирования структуры базы данных
- Б) Метод обработки текстовой информации.
- В) Алгоритм, разбивающий пространство решений на подмножества и принимающий решения на основе признаков.
- Г) Метод случайного выбора данных.

Ответ:

Задание 5.2.

Что такое «кросс-валидация» в машинном обучении?

- А) Процесс разделения данных на обучающую и тестовую выборки.
- Б) Метод оценки производительности модели на основе независимых данных.
- В) Процесс повторного обучения модели на тех же данных.
- Г) Метод оценки производительности модели с использованием разных подмножеств данных.

Ответ:

Задание 6.2.

Что функция потерь в модели машинного обучения?

Ответ:

Задание 7.2.

Какова основная цель алгоритма кластеризации k-средних?

Ответ:

Задание 8.2.

Какой метод машинного обучения позволяет моделям принимать решения на основе прошлого опыта и обратной связи от окружающей среды?

Ответ:

Задание 9.2.

Приведите примеры и охарактеризуйте не менее трех метрик, которые используются для оценки качества моделей машинного обучения.

Ответ:

Задание 10.2.

Сравните две метрики оценки моделей: среднеквадратическую ошибку (Mean Squared Error, MSE) и среднюю ошибку по модулю (Mean Absolute Error, MAE).

Ответ:

Задание 11.2.

Какая метрика обычно используется для оценки качества моделей классификации?

- А) R2-score
- Б) MAE
- В) Accuracy
- Г) F1-score

Ответ:

Задание 12.2.

Какая из представленных ниже задач является задачей машинного обучения?

- А) Задача, в которой компьютер обучается на данных для выявления закономерностей и принятия решений без явного программирования.
- Б) Задача, в которой необходимо найти оптимальное решение с помощью математических методов.
- В) Задача, заключающаяся в разработке интерфейсов для управления роботами.
- Г) Задача, связанная с созданием новых алгоритмов для повышения производительности компьютеров.

Ответ:

Задание 13.2.

Какие задачи, из ниже перечисленных, являются задачами ранжирования?

- А) обнаружение спама
- Б) задачи поискового вывода
- В) определение наиболее целесообразного способа лечения;

Ответ:

Задание 14.2.

Какой тип экспериментального исследования имеет цель - понимание, на что влияют параметры метода обучения?

- А) исследование задач ранжирования
- Б) исследование задач классификации
- В) исследование на модельных данных

Ответ:

Задание 15.2.

Что называется переобучением?

- А) описание искомого алгоритма как суперпозиции некоторых элементарных функций
- Б) поиск преобразования исходящего пространства признаков в новое пространство существенно меньшей размерности
- В) когда по мере увеличения числа используемых признаков средняя ошибка на обучающей выборке монотонно убывает
- Г) когда средняя оценка на независимых контрольных данных сначала уменьшается, затем проходит через точку минимума и далее только возрастает

Ответ:

Задание 16.2.

Какой тип архитектуры нейронной сети используется для данных последовательностей, таких как обработка естественного языка и анализ временных рядов?

Ответ:

Задание 17.2.

Какой метод используется для уменьшения дисперсии модели машинного обучения путем объединения прогнозов из нескольких моделей?

Ответ:

Задание 18.2.

Перечислите основные этапы разработки модели машинного обучения:

Ответ:

Задание 19.2.

Поясните, в чём заключается разница между машинным обучением с учителем и без учителя?

Ответ:

Задание 20.2.

Какие три основных вида слоев обычно используются в сверточных нейронных сетях? Как они обычно сочетаются?

Ответ:

Ключи к оцениванию заданий

Вариант 1

№ задания	Верный ответ	Критерии оценивания
1.1	Б	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
2.1	А	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
3.1	Б	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
4.1	А	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
5.1	Г	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
6.1	Машинное обучение	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
7.1	Ансамблевый метод — это метод машинного обучения, при котором несколько моделей обучаются для решения одной и той же проблемы и объединяются для получения лучших результатов.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
8.1	Данными в машинном обучении называют матрицы, объекты и признаки.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
9.1	Переобучение – это ситуация, когда модель машинного обучения хорошо работает с данными из обучающей выборки, но плохо справляется с новыми данными. Чтобы избежать переобучения, можно предпринять следующие шаги: 1. Собрать больше данных для обучения модели. Чем больше данных, тем меньше вероятность переобучения. 2. Использовать более простые модели, которые менее склонны к переобучению. Например, линейные модели вместо сложных нейронных сетей. 3. Провести кросс-валидацию, чтобы убедиться, что модель не переобучается на обучающих данных. 4. Уменьшить размер модели, удалив ненужные параметры или слои. Это может помочь предотвратить переобучение. 5. Добавить регуляризацию, которая штрафует модель за слишком сложные решения и помогает избежать переобучения.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
10.1	При выборе модели машинного обучения необходимо учитывать следующие факторы: Тип задачи. В зависимости от того, что именно вы хотите сделать (классификация, регрессия, кластеризация и т. д.), вам потребуется выбрать соответствующую модель. Характеристики данных. Размер, структура, тип данных (числовые, категориальные, текстовые) — всё это влияет на выбор модели. Некоторые модели лучше работают с определёнными типами данных.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие

	Сложность модели. Более сложные модели могут давать более точные результаты, но они также требуют больше времени для обучения и могут быть подвержены переобучению. Скорость работы. Если модель должна работать в режиме реального времени, то важно выбрать модель, которая быстро обучается и делает прогнозы. Бюджет и ресурсы. Некоторые модели требуют больших вычислительных ресурсов или специализированного оборудования. Учитывайте свои возможности при выборе модели.	
11.1	Б	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
12.1	Г	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
13.1	А, Б, В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
14.1	Б	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
15.1	Б	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
16.1	Применения машинного обучения включают распознавание изображений, обработку естественного языка, предсказание спроса в бизнесе и другое.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
17.1	Обучение модели.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
18.1	Рекомендательные системы используют алгоритмы машинного обучения для анализа предпочтений пользователей и предлагают им соответствующие товары, услуги или контент.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
19.1	Основные типы задач включают: Регрессию (прогнозирование числовых значений); Классификацию (отнесение объектов к определённым категориям); Кластеризацию (группировку объектов по сходству); Ассоциативные правила (поиск закономерностей между объектами).	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
20.1	Тестовый набор данных используется для оценки качества модели после ее обучения. Валидационный набор используется в процессе обучения для выбора гиперпараметров и предотвращения переобучения на тренировочном наборе.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие

Вариант 2

№ задания	Верный ответ	Критерии оценивания
1.2	В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
2.2	В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
3.2	В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
4.2	В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
5.2	Г	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
6.2	Функция потерь в модели машинного обучения измеряет разность между прогнозируемыми и фактическими значениями.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
7.2	Минимизация дисперсии внутри кластера.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
8.2	Обучение с подкреплением.	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
9.2	Для оценки качества моделей машинного обучения используются следующие метрики: Точность (Assigasy) - доля правильно классифицированных объектов среди всех объектов. Точность (Precision) и полнота (Recall) - метрики, которые используются для оценки качества классификации. Коэффициент корреляции Пирсона - показывает степень линейной зависимости между предсказанными и истинными значениями. Среднеквадратичная ошибка (Mean Squared Error, MSE) - сумма квадратов разностей между предсказанными и истинными значениями, деленная на количество объектов. Площадь под ROC-кривой (Area Under the ROC Curve, AUC-ROC) — количественная характеристика качества бинарной классификации.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
10.2	Среднеквадратичная ошибка (Mean Squared Error, MSE) используется чаще, поскольку она "подсвечивает" большие ошибки. Поскольку производная от x^2 равна $2x$, чем больше x , тем больше разность между x и $x-1$. Однако, иногда выбирают Среднюю ошибку по модулю (Mean Absolute Error, MAE), поскольку она выдает результат, который проще интерпретировать. Таким образом, MSE может быть лучше, если вам просто нужно сравнивать модели друг с другом, но не нужно интерпретировать саму метрику, но если нужно интерпретировать результат метрики, лучше использовать MAE (например, модель в среднем ошибается на \$4).	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
11.2	В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие

12.2	А	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
13.2	Б,В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
14.2	В	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
15.2	Г	1 балл – полное совпадение с верным ответом 0 баллов – неверный ответ или его отсутствие
16.2	Рекуррентная нейронная сеть (RNN)	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
17.2	Бэггинг	2 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
18.2	Основные этапы разработки модели машинного обучения: сбор данных, предобработка данных, выбор модели, обучение модели, тестирование модели, оценка качества модели.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
19.2	Машинное обучение с учителем – это тип машинного обучения, при котором модель обучается на размеченных данных. То есть для каждого примера в наборе данных уже известно правильное решение или ответ. Модель учится сопоставлять входные данные с правильными ответами. Машинное обучение без учителя – это метод машинного обучения, который работает с неструктурированными данными и не требует заранее подготовленных ответов. Цель такого обучения найти скрытые паттерны, структуры или группировки в данных. Модель сама пытается выявить закономерности и зависимости между объектами.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие
20.2	Три основных вида слоев, используемых в сверточных нейронных сетях – это: Сверточный слой: слой, выполняющий операцию свертки, которая создает несколько окон-картинок, обобщая изображение. Слой активации (обычно ReLU): привносит в сеть нелинейность и приводит все негативные пиксели к нулю. Вывод превращается в исправленную карту признаков. Слой группировки (pooling): операция сокращения, которая сокращает размерность карты признаков. Сверточная нейронная сеть обычно состоит из нескольких последовательностей сверточного, активационного и группирующего слоев. За всем этим могут следовать один или два полносвязных слоя или слоя исключения (dropout) для дальнейшего обобщения, и последним идет полносвязный слой.	3 балла – полное совпадение с верным ответом 1 балл – правильный ответ, без обоснования 0 баллов – неверный ответ или его отсутствие