Приложение 4

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»

Институт экономики и предпринимательства

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Введение в анализ данных и искусственный интеллект

Уровень высшего образования
Специалитет

Направление подготовки / специальность
38.05.02 - Таможенное дело

Направленность образовательной программы
Таможенные операции и таможенный контроль

Форма обучения
очная, заочная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина ФТД.03 Введение в анализ данных и искусственный интеллект является факультативом в образовательной программе.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	**Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции**		**Наименование оценочного средства**	
	Индикатор достижения компетенции (код, содержание индикатора)	**Результаты обучения по дисциплине**	**Для текущего контроля успеваемости**	**Для промежуточной аттестации**
УК-1: Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	*УК-1.1: Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними* *УК-1.2: Разрабатывает и содержательно аргументирует стратегию решения проблемной ситуации на основе системного подхода*	*УК-1.1:* *Знает основные типы данных для по-строения моделей* *Умеет оценивать потребность в данных* *УК-1.2:* *Имеет навыки постановки задач машинного обучения*	*Собеседование*	*Зачёт:* *Проект* *Тест*

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная	**заочная**
Общая трудоемкость, з.е.	**1**	**1**
Часов по учебному плану	**36**	**36**
в том числе		
аудиторные занятия (контактная работа):		
- занятия лекционного типа	**8**	**4**
- занятия семинарского типа (практические занятия / лабораторные работы)	**8**	**0**
- КСР	**1**	**1**
самостоятельная работа	**19**	**27**
Промежуточная аттестация	**0** **Зачёт**	**4** **Зачёт**

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)		в том числе							
			Контактная работа (работа во взаимодействии с преподавателем), часы из них						Самостоятельная работа обучающегося, часы	
			Занятия лекционного типа		Занятия семинарского типа (практические занятия/лабораторные работы), часы		Всего			
	очно	заочно	очно	заочно	очно	заочно	очно	заочно	очно	заочно
Тема 1. Понятие машинного обучения: общая характеристика, задачи, виды	7	5.5	1	0.5	1	0	2	0.5	5	5
Тема 2. Машинное обучение с учителем: задачи классификации и регрессии	9	6	2	1	2	0	4	1	5	5
Тема 3. Машинное обучение без учителя: понижение размерности и задачи кластеризации	7	6	2	1	2	0	4	1	3	5
Тема. 4. Анализ и кластеризация текстов	7	6	2	1	2	0	4	1	3	5
Тема 5. Введение в нейронные сети	5	7.5	1	0.5	1	0	2	0.5	3	7
Аттестация	0	4								
КСР	1	1					1	1		
Итого	36	36	8	4	8	0	17	5	19	27

Содержание разделов и тем дисциплины

Тема 1. Понятие машинного обучения: общая характеристика, задачи, виды Что такое модели с учителем. Типы данных для построения регрессионных моделей. Примеры предобработки и визуализации данных для построения регрессии
Библиотеки для построения регрессионных моделей. Построение регрессионных моделей и оценка их валидности. Метрики качества моделей. Визуализация результатов построения регрессионных моделей Регуляризация регрессионных моделей – назначение и примеры применения
Тема 2. Машинное обучение с учителем: задачи классификации и регрессии Типы данных для построения логистической регрессии. Примеры предобработки и визуализации количественных и качественных переменных.
Библиотека для построения логистической регрессии. Балансировка классов с применением библиотеки SMOTE. Построение модели логистической регрессии.
Оценка качества модели логистической регрессии. Оценка качества классификатора. Roc -кривые Построение модели с применением алгоритма случайный лес – используемые библиотеки, трактовка результатов построения модели
Тема 3. Машинное обучение без учителя: понижение размерности и задачи кластеризации Задачи без учителя – общая характеристика. Виды задач без учителя. Типы данных для реализации задач. Примеры практических задач.
Задачи понижения размерности – основное назначение и геометрическая интерпретация. Задачи кластеризации – назначение, K-means++ – определение, библиотеки для реализации, реализация метода K-means++, визуализация результатов, выводы примеры прикладных задач.
Тема 4. Анализ и кластеризация текстов. Назначение анализа текстов. Понятие обработки текстов на естественном языке. Этапы обработки текстов. Кластеризация текстов -библиотеки, практическая реализация, выводы. Модели группы Word2vec
Тема 5. Введение в нейронные сети Основы обучения нейронных сетей. Архитектуры нейронных сетей. Прикладные задачи, решаемые с применением рейнонных сетей

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для обеспечения самостоятельной работы обучающихся используются:
- электронный курс "Машинное обучение с использованием больших данных" (https://e-learning.unn.ru/course/view.php?id=4500).

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции УК-1:

1 Что такое модели с учителем. Типы данных для построения регрессионных моделей. Примеры предобработки и визуализации данных для построения регрессии

2 Библиотеки для построения регрессионных моделей. Построение регрессионных моделей и оценка их валидности. Метрики качества моделей. Визуализация результатов построения регрессионных моделей

3 Регуляризация регрессионных моделей – назначение и примеры применения

4 Типы данных для построения логистической регрессии. Примеры предобработки и визуализации количественных и качественных переменных.

Библиотека для построения логистической регрессии. Балансировка классов с применением библиотеки SMOTE. Построение модели логистической регрессии.

5 Оценка качества модели логистической регрессии. Оценка качества классификатора. Roc -кривые

6 Понятие дерева решений. Типы данных для построения дерева решений. Решение задач классификации и регрессии с применением дерева решений. Параметры модели дерева решений

7 Построение дерева решений – используемые библиотеки, трактовка результатов построения модели

8 Понятие ансамблевого алгоритма решений. Типы данных для применения алгоритма Случайный лес. . Решение задач классификации и регрессии с применением алгоритма случайный лес. Параметры полученной модели и как их можно регулировать

9 Построение модели с применением алгоритма случайный лес – используемые библиотеки, трактовка результатов построения модели

10 Задачи без учителя – общая характеристика. Виды задач без учителя. Типы данных для реализации задач. Примеры практических задач

11 Задачи понижения размерности -основное назначение, способы реализации и геометрическая интерпретация

12 Метод главных компонент – библиотеки для реализации, реализация метода главных компонент, визуализация результатов, выводы

13 Метод сингулярного разложения – библиотеки для реализации, реализация метода сингулярного разложения, визуализация результатов, выводы

14 Задачи кластеризации – назначение, примеры прикладных задач

15 K-means – определение, библиотеки для реализации, реализация метода K-means, визуализация результатов, выводы

16 K-means++ – определение, библиотеки для реализации, реализация метода K-means++, визуализация результатов, выводы

17 Назначение анализа текстов. Понятие обработки текстов на естественном языке. Этапы обработки текстов

18 Предварительная обработка данных: назначение, используемые библиотеки. Лемматизация и стемминг

19 Меры сходства текстов, библиотеки для анализа текстов

20 Кластеризация текстов -библиотеки, практическая реализация, выводы

Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
зачтено	Ответы на вопросы даются без грубых ошибок
не зачтено	Ответы на вопросы даются с грубыми ошибками или студент отказывается от ответа

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компет	**плохо**	**неудовлетворительно**	**удовлетворительно**	**хорошо**	**очень хорошо**	**отлично**	**превосходно**
	не зачтено		**зачтено**				

енций)							
Знания	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
Умения	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продемонстрированы все основные умения. Решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Навыки	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка		**Уровень подготовки**
зачтено	**превосходно**	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы

		одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	**неудовлетворительно**	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Проект) для оценки сформированности компетенции УК-1

Тема проекта выбирается обучающимися самостоятельно в зависимости от направления подготовки, а также решаемых научных и практических задач.

Примеры тем проекта

- Диагностика взаимосвязи уровня экономической безопасности и отдельных показателей качества жизни населения регионов России или
- Анализ отдельных показателей качества жизни населения регионов России

Тема проекта выбирается обучающимися самостоятельно. Примеры тем проекта

- Анализ цен на недвижимость
- Анализ рейтингов регионов по показателям инвестиционного риска, инвестиционного потенциала, ESG рейтинга
- Анализ экологических затрат предприятия
- Анализ затрат предприятий на информационную безопасность
- Анализ расходов на рекламу в организации
- Анализ рейтинга фильмов на сайте «Кинопоиск»
- Анализ уровня финансовой грамотности населения

Проект должен включать:

- Постановку проблемы.
- Формулировку целей и задач проекта.
- Подбор данных для анализа
- Статистический анализ и визуализацию данных.

- Обоснование выбора моделей, построение моделей, который описывали бы взаимосвязи между показателя
- Оценку качества моделей
- Прогнозирование на основе модели
- Обоснование полученных результатов
- Выводы

Проект должен в обязательно порядке содержать файлы с набором команд для построения моделей (код). Комментарии в коде обязательны.

Критерии оценивания (оценочное средство - Проект)

Оценка	Критерии оценивания
зачтено	Проект выполнен без грубых ошибок
не зачтено	Проект выполнен с грубыми ошибками

5.3.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции УК-1

Вопрос 1. Модель линейной регрессии предполагает «линейную связь между входными переменными и единственной выходной переменной». В чем смысл этого предположения?

(A) Выходная переменная не может быть вычислена на основе линейной комбинации входных переменных

(B) Выходная переменная может быть вычисляется из линейной комбинации входных переменных

(C) Входные переменные могут быть вычислены из линейной комбинации выходных переменных

Вопрос 2: В простой задаче линейной регрессии с одной входной переменной (x) и одной выходной переменной (y) линейное уравнение будет иметь вид y = ax + b; где a и b _____ и _____ соответственно. (выберите два)

(A) коэффициент смещения, коэффициент характеристики

(B) коэффициент характеристики, коэффициент смещения

(C) наклон, точка пересечения по оси y

(D) пересечение оси y, наклон

Вопрос 3. Для линии регрессии через данные вертикальное расстояние от каждой точки данных до линии регрессии называется невязкой. (i) возвести остаток в квадрат и (ii) суммировать все возведенные в квадрат ошибки. Это количество, которое обычные методы наименьших квадратов стремятся _____?

(A) минимизировать

(B) максимизировать

(C) увеличить

(D) Ни один из этих

Вопрос 4. Для модели линейной регрессии начните со случайных значений для каждого коэффициента. Сумма квадратов ошибок вычисляется для каждой пары входных и выходных значений. Скорость обучения используется как масштабный коэффициент, и коэффициенты обновляются в направлении минимизации ошибки. Процесс повторяется до тех пор, пока не будет достигнута минимальная квадратичная ошибка суммы или пока не станет возможным дальнейшее улучшение. Этот метод называется _____?

(A) Градиентный спуск

(B) Обычные наименьшие квадраты

(C) Гомоскедастичность

(D) Регуляризация

Вопрос 5. Какой параметр определяет размер шага улучшения, выполняемого на каждой итерации градиентного спуска?

(A) Скорость обучения

(B) эпоха

(C) размер пакета

(D) параметр регуляризации

Вопрос 6. Одно из основных предположений линейной регрессии: когда дисперсия вокруг линии регрессии одинакова для всех значений переменной-предиктора, называется _____?

(A) Регуляризация L1

(B) Регрессия Лассо

(C) Гомоскедастичность

(D) Гетероскедастичность

Вопрос 7. Для модели линейной регрессии мы выбираем коэффициенты и член смещения, минимизируя _____.

(A) Функция потерь

(B) Функция ошибок

(C) Функция затрат

(D) Все вышеперечисленное

Вопрос 8: Какое из них является правильным предположением линейной регрессии?

(A) Линейная регрессия предполагает, что входные и выходные переменные не зашумлены

(B) Линейная регрессия будет превосходить ваши данные, если у вас есть сильно коррелированные входные переменные

(C) Остатки (истинное целевое значение - прогнозируемое целевое значение) данных нормально распределены и независимы друг от друга

Вопрос 9. Какой метод может найти коэффициенты в модели линейной регрессии?

(A) Обычные методы наименьших квадратов

(B) Градиентный спуск

(C) Регуляризация

(D) Все вышеперечисленное

Вопрос 10: В чем заключается недостаток линейной регрессии?

(A) Предположение о линейности между зависимой переменной и независимыми переменными. В реальном мире данные не всегда линейно разделимы

(B) Линейная регрессия очень чувствительна к выбросам

(C) Перед применением линейной регрессии мультиколлинеарность должна быть удалена, поскольку она предполагает отсутствие связи среди независимых переменных.

(D) Все вышеперечисленное

Вопрос 11. Логистическая регрессия используется для ___?

(A) классификация

(B) регрессия

(C) кластеризация

(D) Все эти

Вопрос 12: логистическая регрессия - это алгоритм машинного обучения, который используется для прогнозирования вероятности ___?

(A) категориальной независимой переменной

(B) категориальной зависимой переменной. < br /> (C) числовая зависимая переменная.

(D) числовая независимая переменная.

Вопрос 13: вы прогнозируете, является ли электронное письмо спамом. Основываясь на характеристиках, вы получили оценку вероятности 0,75. Что означает эта предполагаемая вероятность? (выберите два)

(A) вероятность того, что письмо будет спамом, составляет 25%

(B) вероятность того, что письмо будет спамом, составляет 75%

(C) вероятность того, что письмо не будет спамом, составляет 75%.

(D) Вероятность того, что письмо не будет спамом, составляет 25%.

Вопрос 14. В модели логистической регрессии граница принятия решения может быть ___.

(A) линейная

(B) нелинейная

(C) обе (A) и (B)

(D) ни один из этих

Вопрос 15. Какова функция затрат логистической регрессии?

(A) Сигмоидная функция

(B) Логистическая функция

(C) и (A), и (B)

(D) ни один из этих

Вопрос 16: Почему функцию стоимости, которая использовалась для линейной регрессии, нельзя использовать для логистической регрессии?

(A) Линейная регрессия использует среднеквадратичную ошибку в качестве функции стоимости. Если это используется для логистической регрессии, то это будет невыпуклая функция своих параметров. Градиентный спуск сведется к глобальному минимуму, только если функция выпуклая.

(B) Линейная регрессия использует среднеквадратичную ошибку в качестве функции стоимости. Если это используется для логистической регрессии, то это будет выпуклая функция от своих параметров. Градиентный спуск сведется к глобальному минимуму, только если функция выпуклая.

(C) Линейная регрессия использует среднеквадратичную ошибку в качестве функции стоимости. Если это используется для логистической регрессии, то это будет невыпуклая функция своих параметров. Градиентный спуск сведется к глобальному минимуму, только если функция невыпуклая.

(D) Линейная регрессия использует среднеквадратичную ошибку в качестве функции стоимости. Если это используется для логистической регрессии, то это будет выпуклая функция от своих параметров. Градиентный спуск сведется к глобальному минимуму только в том случае, если функция невыпуклая.

Вопрос 17. Вы прогнозируете, является ли электронное письмо спамом. Основываясь на характеристиках, вы получили оценку вероятности 0,75. Что означает эта предполагаемая вероятность? Пороговое значение для различения классов составляет 0,5.

(A) Электронное письмо не является спамом

(B) Электронное письмо является спамом

(C) Не могу определить

(D) и (A), и (B)

Вопрос 18: какова гипотеза логистической регрессии?

(A) для ограничения функции стоимости от 0 до 1

(B) для ограничения функции стоимости от -1 до 1

(C) для ограничения функции стоимости между -infinity и + infinity

(D) для ограничения функции стоимости от 0 до + бесконечности

Вопрос 19: какой из них неверен?

(A) Если мы возьмем взвешенную сумму входных данных в качестве выходных данных, как в случае линейной регрессии, значение может быть больше 1, но мы нужно значение от 0 до 1. Вот почему линейную регрессию нельзя использовать для задач классификации.

(B) Логистическая регрессия - это обобщенная линейная регрессия в том смысле, что мы не выводим взвешенную сумму входных данных напрямую, а передаем ее через функцию, которая может отображать любое реальное значение от 0 до 1.

(C) Значение сигмовидной функции всегда находится между 0 и 1.

(D) Логистическая регрессия используется для определения значения непрерывной зависимой переменной.

Вопрос 20. В чем заключается проблема проклятия размерности в методе k ближайших соседей?

(A)Увеличение размерности приводит к увеличению времени поиска ближайших соседей.

(B)Проблема проклятия размерности не относится к методу k ближайших соседей.

(C) Увеличение размерности приводит к увеличению ошибки модели.

(D) Увеличение размерности приводит к увеличению числа примеров в обучающем наборе.

Вопрос 21. Какова максимальная точность, которую можно достичь с помощью метода k ближайших соседей?

(A)90%.

(B)Точность не ограничена.

(C) 70%.

(D)50%.

Вопрос 22. Что такое метод k ближайших соседей?

(A) Метод машинного обучения для классификации и регрессии данных.

(B) Метод градиентного спуска.

(B) Метод вычисления статистических показателей.

(D) Метод оптимизации функций.

Вопрос 23. Каким образом выбираются веса для взвешенного метода k ближайших соседей?

(A)Используется расстояние до каждого ближайшего соседа и их порядковый номер.

(B)Используется расстояние до каждого ближайшего соседа и их классы.

(C) Задается вручную.

(D) Используется расстояние до каждого ближайшего соседа.

Вопрос 24. Каким образом можно использовать метод k ближайших соседей для задачи классификации?

(A)Выбрать k на основе минимального расстояния до k ближайших соседей и отнести объект к классу, который наиболее часто встречается среди k ближайших соседей.

(B) Выбрать k на основе максимального расстояния до k ближайших соседей и отнести объект к классу, который наиболее часто встречается среди k ближайших соседей.

(C)Нельзя использовать метод k ближайших соседей для задачи классификации.

(D)Использовать метод k ближайших соседей только для задач регрессии.

Вопрос 25. Как предпочтительно выбирать значение k в методе k ближайших соседей?

(A)Значение k выбирается на основе экспертного мнения.

(B)Значение k выбирается случайным образом.

(C)Значение k выбирается на основе кросс-валидации.

(D) Значение k не имеет значения.

Вопрос 26. Каким образом можно уменьшить влияние выбросов непосредственно при обучении модели машинного обучения методом k ближайших соседей?

(A)Использовать алгоритмы кластеризации для определения выбросов и исключения их из обучающего набора позволяет.

(B)Уменьшить значение k.

(C)Использовать взвешенный метод k ближайших соседей, где вес каждого ближайшего соседа зависит от расстояния до нового примера.

(D)Использовать метрики расстояния, устойчивые к выбросам, например, манхэттенское расстояние.

Вопрос 28. Что такое евклидово расстояние в методе k ближайших соседей?

(A) Математическое ожидание расстояния.

(B)Расстояние между двумя линиями.

(C)Мера сходства между двумя примерами.

(D)Расстояние между точками на карте.

Вопрос 29. Какой алгоритм используется для поиска k ближайших соседей?

(A)Жадный алгоритм.

(B)Алгоритм поиска ближайших соседей.

(C)Алгоритм Беллмана-Форда.

Вопрос 30. Что такое дерево решений?

(A)Функция активации для нейронной сети

(B)Алгоритм машинного обучения

(C)Оптимизационный алгоритм

(B)Метод решения математических уравнений

Вопрос (31). Какой критерий используется для разбиения узла дерева решений?

(A) Наименьшее увеличение неопределенности (smallest increase in uncertainty)

(B)Наибольшее уменьшение неопределенности (largest reduction in uncertainty)

(C)Наибольшее увеличение неопределенности (largest increase in uncertainty)

(D)Наименьшее уменьшение неопределенности (smallest reduction in uncertainty)

Вопрос 32. Каким образом дерево решений выбирает наилучший признак для разделения?

(A)Методом случайного выбора признаков

(B)Методом наименьших квадратов

(C)Методом информационного выигрыша

(D)Методом градиентного спуска

Вопрос 33. Какой из следующих методов может быть использован для улучшения работы дерева решений на несбалансированных данных?

(A)Увеличение размера обучающей выборки

(B)Использование ансамблевых методов, таких как случайный лес (random forest)

(C)Использование алгоритмов обрезки дерева

(D)Использование взвешенных функций ошибки

Вопрос 34. Способно ли дерево решений обрабатывать отсутствующие значения (missing values)?

(A) Нет

(B)Да

(C)Только если пропущенное значение это категориальные данные

(D)Только если пропущенное значение это численные данные

Вопрос 35. Каким образом дерево решений может быть визуализировано?

(A) С помощью круговой диаграммы

(B) С помощью диаграммы дерева решений

(D) С помощью графика рассеяния (scatter plot)

(D) С помощью гистограммы

Вопрос 36. Каким образом дерево решений может использоваться для отбора признаков (feature selection)?

(A)Путем случайного выбора признаков

(B)Путем ручного выбора наиболее важных признаков

(C) Путем увеличения количества признаков

(D) Путем исключения признаков с низким информационным выигрышем

Вопрос 37. Какие типы задач можно решить с помощью дерева решений?

(A)Кластеризация и ассоциативные правила

(B)Классификация и регрессия

(C)Прогнозирование временных рядов и нейронные сети

(D)Распознавание образов и генетические алгоритмы

Вопрос 38. Каким образом решают проблему переобучения при обучении деревьев решений?

(A)Путем уменьшения количества объектов в обучающей выборке

(B)Путем уменьшения количества признаков

(C)Путем использования алгоритмов обрезки дерева или ограничения глубины дерева

(D)Путем увеличения глубины дерева

Вопрос 39. Что такое лист дерева решений?

(A)Узел дерева решений, который не имеет потомков

(B)Корень дерева решений

(C)Узел дерева решений, который имеет потомков

(D)Узел дерева решений, который имеет только одного потомка

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	60 % и более
не зачтено	менее 60%

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Козлов Андрей Юрьевич. Статистический анализ данных в MS Excel : Учебник / Пензенский государственный университет; Национальный исследовательский университет "Высшая школа экономики"; Военная академия материально-технического обеспечения им. генерала армии А.В. Хрулёва, ф-л г. Пенза. - 1. - Москва : ООО "Научно-издательский центр ИНФРА-М", 2021. - 320 с. - ВО - Бакалавриат. - ISBN 978-5-16-004579-5. - ISBN 978-5-16-101024-2., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=770671&idb=0.
2. Кулаичев Алексей Павлович. Методы и средства комплексного статистического анализа данных : Учебное пособие / Московский государственный университет им. М.В. Ломоносова, биологический факультет. - 5. - Москва : ООО "Научно-издательский центр ИНФРА-М", 2022. - 484 с. - ВО - Бакалавриат. - ISBN 978-5-16-012834-4. - ISBN 978-5-16-103357-9., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=771058&idb=0.
3. Дайитбегов Дайитбег Магамедович. Компьютерные технологии анализа данных в эконометрике : Монография / Финансовый университет при Правительстве Российской Федерации. - 3-е изд. ; доп. - Москва : Вузовский учебник, 2018. - 587 с. - Дополнительное профессиональное образование. - ISBN 978-5-9558-0275-6. - ISBN 978-5-16-500249-6. - ISBN 978-5-16-006145-0., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=742168&idb=0.

Дополнительная литература:

1. Лемешко Борис Юрьевич. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : Монография / Новосибирский государственный технический университет. - Москва : ООО "Научно-издательский центр ИНФРА-М", 2015. - 890 с. - Дополнительное профессиональное образование. - ISBN 978-5-16-103267-1., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=594609&idb=0.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

Программное обеспечение

№ п/п Наименование Условия доступа
1. Windows Professional 8.1 Russian Из внутренней сети университета (договор)
2. MS Office 2007 Prof+ Из внутренней сети университета (договор)
3. Среда Anaconda Navigator Программный продукт свободного доступа
4. Jupyter Notebook - командная оболочка для интерактивных вы-числений Программный продукт свободного доступа

5. Google Colab Программный продукт свободного доступа

Интернет-ресурсы

№ п/п	Наименование	Адрес web-страницы
1.	GitHub - веб-сервис для хостинга IT-проектов	https://github.com
2.	Kaggle – сеть специалистов по обработке данных	https://www.kaggle.com/datasets
3.	Habr – ресурс для IT специалистов	https://habr.com/ru/all/

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 38.05.02 - Таможенное дело.

Автор(ы): Граница Юлия Валентиновна, кандидат экономических наук, доцент.

Заведующий кафедрой: Болдыревский Павел Борисович, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 12.12.2023, протокол № 6.