

МИНОБРНАУКИ РОССИИ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО ВГУ)

УТВЕРЖДАЮ
Заведующий кафедрой
Математических методов исследования операций

Азарнова Т.В.
21.06.2021



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Б1.В.03 Машинное обучение

- 1. Шифр и наименование направления подготовки/специальности:**
01.03. 02 Прикладная математика и информатика
- 2. Профиль подготовки/специализации:** все профили
- 3. Квалификация (степень) выпускника:** бакалавр
- 4. Форма образования:** очная
- 5. Кафедра, отвечающая за реализацию дисциплины:** математических методов исследования операций
- 6. Составители программы:** Каширина Ирина Леонидовна, доктор техн. наук, профессор кафедры математических методов исследования операций
- 7. Рекомендована:** НМС факультета прикладной математики, информатики и механики, протокол №9 от 23.05.2020. НМС факультета Прикладной математики, информатики и механики, протокол №10 от 15.06.2021
- 8. Учебный год:** 2021/2022 **Семестр(-ы):** 6

9. Цели и задачи учебной дисциплины:

Целью курса является изучить теоретическую базу в сфере интеллектуальных методов анализа и приобрести разнообразные практические навыки, которые позволят слушателям непосредственно участвовать в реализации программ и проектов, связанных с информационными технологиями поддержки принятия решений.

Задачи курса:

- ознакомить студентов с основными понятиями машинного обучения;
- ознакомить студентов с основными классами и принципами обучения нейронных сетей;
- дать представление об алгоритмах построения деревьев решений;
- изучить методы классификации и кластеризации
- сформировать у студентов практические навыки по использованию моделей и методов машинного обучения для решения прикладных задач.

10. Место учебной дисциплины в структуре ООП:

Дисциплина относится к дисциплинам по выбору вариативной части Блока 1. Для изучения курса необходимы базовые знания теории вероятностей и математической статистики.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ОПК-2	Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач	ОПК-2.2	Анализирует требования к решению прикладной задачи, выделяя основные направления адаптации и модификации методов ее решения; реализует численные методы и алгоритмы в подходящей программной среде для проведения вычислительного эксперимента.	знать: - возможности актуальных алгоритмов машинного обучения, которые широко используются на практике, основные сферы их применения; уметь: – применять методы машинного обучения при решении задач в различных прикладных областях; – использовать библиотеки языка Python для построения моделей машинного обучения; владеть (иметь навык(и)): - технологиями разработки алгоритмов и программными системами машинного обучения.
ПКВ-2	Способен проводить отдельные виды исследований в рамках поставленных задач по стандартным методикам	ПКВ-2.2	Проводит эксперимент в соответствии с установленными полномочиями, составляет его описание и формулирует выводы	знать: - основные задачи и методы машинного обучения. уметь: - формулировать задачи машинного обучения, выбирать и разрабатывать адекватные алгоритмы их решения, оценивать качество получаемых решений. владеть (иметь навык(и)): - проведения вычислительных экспериментов для оценки качества получаемых решений; - интерпретации полученных результатов анализа в терминах прикладной области с целью получения новых знаний и выводов;

12. Объем дисциплины в зачетных единицах/час — 3/108

Форма промежуточной аттестации *зачет*

13. Виды учебной работы

Вид учебной работы	Трудоемкость (часы)	
	Всего	По семестрам
		6 сем.
Аудиторные занятия	64	64
в том числе: лекции	32	32
практические		
лабораторные	32	32
Самостоятельная работа	44	44
Итого:	108	108
Форма промежуточной аттестации	<i>зачет</i>	<i>зачет</i>

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
1. Лекции			
1.1	Задачи машинного обучения.	Классификация задач машинного обучения. От данных к решениям. Сопоставление и сравнение понятий "информация", "данные", "знание". Сферы применения машинного обучения	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
1.2	Классификация и кластеризация	Задача классификации. Процесс классификации Методы, применяемые для решения задач классификации Точность классификации: оценка уровня ошибок. Оценивание классификационных методов. Задача кластеризации. Оценка качества кластеризации. Процесс кластеризации. Применение кластерного анализа. Кластерный анализ в маркетинговых исследованиях. Практика применения кластерного анализа в маркетинговых исследованиях.	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
1.3	Прогнозирование и визуализация	Задача прогнозирования Сравнение задач прогнозирования и классификации Прогнозирование и временные ряды Тренд, сезонность и цикл Точность прогноза Виды прогнозов Методы прогнозирования Задача визуализации	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579

			d=3579
1.4.	Методы классификации и прогнозирования. Ансамблирование.	Деревья решений. Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация. Нейронные сети. Построение ансамблей классификаторов: случайный лес, бустинг, бэггинг.	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
1.5	Методы кластерного анализа. Отбор значимых признаков.	Иерархические методы (агломеративные и дивизимные, дендрограмма). Итеративные методы (к-средних) . Метрики расстояния между объектами и кластерами. Методы отбора признаков (метод главных компонент)	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
1.6	Методы поиска ассоциативных правил	Часто встречающиеся приложения с применением ассоциативных правил: Введение в ассоциативные правила. Часто встречающиеся шаблоны или образцы. Характеристики ассоциативных правил. Границы поддержки и достоверности ассоциативного правила. Методы поиска ассоциативных правил Разновидности алгоритма Apriori .	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
2. Лабораторные работы			
2.1	Обзор основных необходимых библиотек языка Python	Библиотека NumPy для оптимизированных вычислений над массивами данных. Введение в массивы библиотеки NumPy. Выполнение вычислений над массивами библиотеки NumPy, универсальные функции Операции над данными в библиотеке Pandas. Обработка отсутствующих данных. Агрегирование и группировка. Визуализация с помощью библиотеки Matplotlib. Линейные графики, диаграммы рассеяния, гистограммы, трехмерные графики. Знакомство с библиотекой машинного обучения Scikit-Learn. Гиперпараметры и проверка качества модели	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
2.2	Построение и отбор признаков	Извлечение признаков (Feature Extraction). Преобразования признаков (Feature transformations): кодирование нечисловых данных, нормировка и калибровка, заполнение пропусков Выбор признаков (Feature selection): статистические подходы, визуализация, отбор с использованием моделей	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
2.3	Классификация и кластеризация: Древоподобные модели: деревья решений, случайный лес	Построение моделей деревьев решений и случайного леса с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели	Курс «Машинное обучение» на портале «Электронный университет ВГУ»

			ВГУ». https://edu.vsu.ru/course/view.php?id=3579
2.4	Классификация и кластеризация: Методы кластеризации	Построение моделей кластеризации с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579
2.5	Методы поиска ассоциативных правил	Построение моделей ассоциативных правил с помощью библиотеки Apriori для заданного набора данных. Анализ качества построенной модели	Курс «Машинное обучение» на портале «Электронный университет ВГУ». https://edu.vsu.ru/course/view.php?id=3579

13.2. Темы (разделы) дисциплины и виды занятий:

№ п/п	Наименование раздела дисциплины	Виды занятий (часов)				Всего
		Лекции		Лабораторные	Самостоятельная работа	
1	Задачи машинного обучения.	4		0	6	10
2	Обзор основных необходимых библиотек языка Python	0		4	6	10
3	Классификация и кластеризация	4		4	6	14
4	Прогнозирование и визуализация	4		4	6	14
5	Методы классификации и прогнозирования. Ансамблирование.	8		8	8	24
6	Методы кластерного анализа. Отбор значимых признаков.	8		8	6	22
7	Методы поиска ассоциативных правил	4		4	6	14
Итого:		32		32	44	

14. Методические указания для обучающихся по освоению дисциплины

(рекомендации обучающимся по освоению дисциплины: работа с конспектами лекций, презентационным материалом, выполнение практических заданий, тестов, заданий текущей аттестации и т.д.)

Работа с конспектами лекций, презентациями, выполнение практических заданий для самостоятельной работы, выполнение лабораторных работ, использование рекомендованной литературы и методических материалов, в том числе размещенных на странице курса «Машинное обучение» на портале «Электронный университет ВГУ» <https://edu.vsu.ru/course/view.php?id=3579>, автор Каширина И.Л.

В рамках общего объема часов, отведенных для изучения дисциплины, предусматривается выполнение следующих видов самостоятельных работ студентов (СРС): изучение теоретического материала, написание программ по темам, изученным на лекционных и практических занятиях.

При использовании дистанционных образовательных технологий и электронного обучения выполнять все указания преподавателей по работе на LMS-платформе, своевременно подключаться к online-занятиям, соблюдать рекомендации по организации самостоятельной работы

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины (список литературы оформляется в соответствии с требованиями ГОСТ и используется общая сквозная нумерация для всех видов источников)

а) основная литература:

а) основная литература:

№ п/п	Источник
1	Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс] : руководство / С. Рашка ; пер. с англ. Логунова А.В.. — Электрон. дан. — Москва : ДМК Пресс, 2017. — 418 с. — Режим доступа: https://e.lanbook.com/book/100905
2	А.Мюллер, С.Гвидо - Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными – 2017 электронный ресурс свободного доступа: https://owlweb.ru/wp-content/uploads/2017/06/a.myuller-s.gvido-vvedenie-v-mashinnoe-obuchenie-s-pomoshhyu-python.-rukovodstvo-dlya-specialistov-po-rabote-s-dannymi-2017.compressed-1.pdf материалы к книге: https://github.com/amueller/introduction_to_ml_with_python
3	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] / П. Флах. — Электрон. дан. — Москва : ДМК Пресс, 2015. — 400 с. — Режим доступа: https://e.lanbook.com/book/69955
4	Козьло, Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Козьло, В. Ричарт ; пер. с англ. Слинкин А. А.. — Электрон. дан. — Москва : ДМК Пресс, 2016. — 302 с. — Режим доступа: https://e.lanbook.com/book/82818

б) дополнительная литература:

№ п/п	Источник
5	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с. Материалы к книге: https://github.com/jakevdp/PythonDataScienceHandbook
6	Астахова И.Ф. , Чулюков В.А., Каширина И. Л. и др. Системы искусственного интеллекта. Практический курс. М. : БИНОМ. Лаборатория знаний, 2008. — 292 с
7	Дюк В.А., Самойленко А.П Data Mining: учебный курс. - СПб.: Питер, 2001
8	Прикладные методы анализа статистических данных [Электронный ресурс] : учебное пособие / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. — Электрон. дан. — М. : Издательский дом Высшей школы экономики, 2012. — 312 с. — Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=65997
9	Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер.с польск.И.Д.Рудинского. [Электронный ресурс] : / Рутковская Д., Пилиньский М., Рутковский Л. — Электрон. дан. — М. : Горячая линия-Телеком, 2013. — 384 с. — Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=11843
10	Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009.
11	Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. Пер. с англ. - СПб.: ООО "Альфа-книга": 2018. - 688 с
12	Бессмертный, И. А. Системы искусственного интеллекта : учебное пособие для вузов / И. А. Бессмертный. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 157 с. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/451721 (дата обращения: 25.12.2020)

в) базы данных, информационно-справочные и поисковые системы:

№ п/п	Источник
-------	----------

13	http://e.lanbook.com/ Электронная библиотечная система «Издательства «Лань»,
14	Курс «Машинное обучение» на портале «Электронный университет ВГУ», автор Каширина И.П. https://edu.vsu.ru/course/view.php?id=3579
15	http://www.lib.vsu.ru Электронная библиотечная система ВГУ
16	http://MachineLearning.ru Ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
17	https://www.kaggle.com/ Kaggle – это платформа для исследователей разных уровней, где они могут опробовать свои модели анализа данных и машинного обучения на серьезных и актуальных задачах.

16. Перечень учебно-методического обеспечения для самостоятельной работы (учебно-методические рекомендации, пособия, задачки, методические указания по выполнению практических (контрольных) работ и др.)

№ п/п	Источник
1	Курс «Машинное обучение» на портале «Электронный университет ВГУ», автор Каширина И.П. https://edu.vsu.ru/course/view.php?id=3579
2	Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017. -336 с.: Материалы к книге: https://github.com/brinkar/real-world-machine-learning
3	UCI Machine Learning Repository — репозиторий наборов данных для выполнения лабораторных работ по курсу Data Mining - http://archive.ics.uci.edu/ml/

17. Информационные технологии, используемые для реализации учебной дисциплины, включая программное обеспечение и информационно-справочные системы (при необходимости)

Python 3 с подключенными библиотеками (дистрибутив Anaconda)

18. Материально-техническое обеспечение дисциплины:

Лекционная аудитория должна быть оснащенной современным компьютером с подключенным к нему проектором с видеотерминала на настенный экран.

Практические и лабораторные занятия должны проводиться в специализированной аудитории, оснащенной современными персональными компьютерами и программным обеспечением в соответствии с тематикой изучаемого материала

Перечень специализированных лабораторий:

Лаборатория машинного обучения (корпус 1, ауд. 407п)

Компьютер в составе (16 шт.): системный блок: процессор Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, оперативная память 16 Гб, SSD 256 Гб, HDD 1Тб, ви-деокарта NVIDIA GeForce GTX 1080 Ti; мо-нитор DELL S2419HN; Компьютер в составе (1 шт.):

системный блок: процессор Intel(R) Core(TM) i7-7800X CPU @ 3.50GHz, опера-тивная память 96 Гб, SSD 1Тб, HDD 4Тб, ви-деокарта NVIDIA GeForce RTX 2080 Ti (2 шт.); монитор DELL S2419HN; Источник бесперебойного питания APC Back-UPS BV1000I-GR, line-interactive, мощ-ность:1000ВА, 600Вт (16 шт.); Источник бесперебойного питания Legrand KEOR LINE RT 1500BA (1 шт.); Коммутатор HP 2530-24G Switch (Managed, 24*10/100/1000 + 4 SFP, 19"); Интерактивная доска SMART SBM685 (87 дюймов, ПО SMART SLS) с пассивным лот-ком; Проектор Vivitek DH758UST (ультракорот-кофокусный, DLP, Full HD 1080p (1920 x 1080) , 3500 ANS, 10000:1, полная поддержка 3D)

Лаборатория искусственного интеллекта (корпус 1, ауд. 124)

Компьютер в составе (17 шт.): системный блок: процессор AMD Ryzen 7 3800X 8-Core Processor, оперативная память 32Гб, HDD 1Тб, SSD 256Гб, видеокарта NVIDIA GeForce GTX 1050; монитор: Dell S2419H; Интерактивная доска SMART SBM685 (87 дюймов); Мультимедиа-проектор Vivitek ультрако-роткофокусный; Источник бесперебойного питания Legrand Keor SPX 1000 BA IEC C13 (16 шт.); Источник бесперебойного питания Legrand Keor Line RT 1000 BA (1 шт.); Коммутатор HP 2530-48G Switch (1 шт.)

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Задачи машинного обучения.	ОПК-2	ОПК-2.2	Тест
2	Обзор основных необходимых библиотек языка Python	ПКВ-2	ПКВ-2.2	Тест
3	Классификация и кластеризация	ОПК-2	ОПК-2.2	Тест
4	Прогнозирование и визуализация	ОПК-2	ОПК-2.2	Задание для лабораторной работы
5	Методы классификации и прогнозирования. Ансамблирование.	ПКВ-2	ПКВ-2.2	Тест
6	Методы кластерного анализа. Отбор значимых признаков.	ПКВ-2	ПКВ-2.2	Тест
7	Методы поиска ассоциативных правил	ОПК-2	ОПК-2.2	Задание для лабораторной работы
Промежуточная аттестация форма контроля – экзамен				<i>Перечень вопросов</i> Практическое задание

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

Тестовые задания, Лабораторные работы, Устный опрос

Тестовые задания

Задание1:

Задачу классификации нельзя решить с помощью...

Вариант 1 алгоритма Apriori

Вариант 2 метода деревьев решений

Вариант 3 нейронных сетей

Задание2:

Регрессия — это...

Вариант 1 это установление зависимости непрерывной выходной переменной от входных переменных

Вариант 2 эта группировка объектов на основе данных, описывающих свойства объектов

Вариант 3 выявление закономерностей между связанными событиями

Задание3:

Основная характеристика задачи бинарной классификации:

Вариант 1 классификация осуществляется по одному признаку

Вариант 2 зависимая переменная может принимать только два значения

Вариант 3 классификация осуществляется по двум признакам

Задание4:

Классификация относится к стратегии:

Вариант 1 обучения без учителя

Вариант 2 обучения с учителем

Вариант 3 оба ответа неверны

Задание5:

Иерархические алгоритмы применяются для решения задач ...

Вариант 1 классификации

Вариант 2 кластеризации

Вариант 3 классификации и кластеризации

Задание6:

Решение задачи прогнозирования ...

Вариант 1 является решением задачи "обучения без учителя"

Вариант 2 возможно без обучающей выборки данных

Вариант 3 требует некоторой обучающей выборки данных

Задание7:

В чем состоит основное сходство задач прогнозирования и классификации?

Вариант 1 оба ответа верны

Вариант 2 при решении обеих задач используется двухэтапный процесс построения модели на основе обучающего набора и ее использования для предсказания неизвестных значений зависимой переменной

Вариант 3 сходство заключается в том, что при решении обеих задач предсказываются числовые значения зависимой переменной

Задание8:

Отличием анализа временных рядов от анализа случайных выборок является:

Вариант 1 оба варианта верны

Вариант 2 их хронологический порядок

Вариант 3 предположение о равных промежутках времени между наблюдениями

Задание9:

Какие из перечисленных технологий относятся к анализу текстов?

Вариант 1 word2vec

Вариант 2 ARIMA

Вариант 3 мешок слов

Вариант 4 tf-idf

Задание10:

Временной ряд — последовательность наблюдаемых значений какого-либо признака,...

Вариант 1 упорядоченных в неслучайные моменты времени

Вариант 2 упорядоченных в случайные моменты времени

Вариант 3 не обязательно упорядоченных, но зафиксированных в неслучайные моменты времени

Задание11:

Выделяют такие этапы построения модели "мешок слов":

Вариант 1 токенизация

Вариант 2 построение словаря

Вариант 3 моделирование тем

Вариант 4 формирование разреженной матрицы

Задание12:

Ошибка обучения нейронной сети называется ...

Вариант 1 разность между желаемым и полученным на выходе сигналами

Вариант 2 переобучение нейронной сети

Вариант 3 целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети

Задание13:

Каждый этап работы алгоритма Apriori состоит из таких шагов:

Вариант 1 подсчет кандидатов

Вариант 2 формирование кандидатов

Вариант 3 кодирование кандидатов

Задание14:

Поддержка ассоциативного правила определяет...

Вариант 1 какая вероятность того, что из события А следует событие В

Вариант 2 процент транзакций, содержащих наборы данных А и В

Вариант 3 количество транзакций, содержащих набор данных А

Задание15:

Изначальная предопределенность классов является характеристикой задачи ...

Вариант 1 классификации

Вариант 2 классификации и кластеризации

Вариант 3 кластеризации

Задание16:

Алгоритм конструирования дерева решений ...

Вариант 1 не требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых

Вариант 2 требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых

Вариант 3 на вход алгоритма можно подавать все существующие атрибуты, алгоритм сам выберет наиболее значимые среди них, и только они будут использованы для построения дерева

Задание17:

Явление переобучения характеризуется ...

Вариант 1 чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению

Вариант 2 возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры модели

Вариант 3 возникновением, в случае слишком долгого обучения, слишком сложной структуры модели

Задание18:

Нейрон имеет аксон, который представляет собой ...

Вариант 1 однонаправленные входные связи, соединенные с выходами других нейронов

Вариант 2 выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов

Вариант 3 один или несколько нейронов, на входы которых подается один и тот же общий сигнал

Задание19:

В качестве функции активации нейрона часто используются функции - ...

Вариант 1 гиперболический синус

Вариант 2 гиперболический тангенс

Вариант 3 логистическая функция

Задание20:

Множество примеров, используемое для конструирования модели, называется...

Вариант 1 обучающим множеством

Вариант 2 тестовым множеством

Вариант 3 валидационным множеством

Задание 21:

Иерархические дивизимные методы характеризуются ...

Вариант 1 сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга

Вариант 2 делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп

Вариант 3 последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров

Задание22:

Преимуществом какой группы методов кластеризации является их наглядность и возможность получить детальное представление о структуре данных

- Вариант 1 иерархические методы
- Вариант 2 оба варианта верны
- Вариант 3 неиерархические методы

Перечень заданий для лабораторных работ

Лабораторная работа № 1

- 1) Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
- 2) Проведите предобработку данных: заполнение пропусков, кодирование, масштабирование
- 3). Обучите, а затем провалидируйте на тестовых данных модель случайного леса
- 4) Вычислите значения метрик: recall, precision, F1-мера, AUC-ROC. Постройте ROC-кривую.

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств:

Теоретические вопросы, практические задания

Контрольно-измерительные материалы промежуточной аттестации включают в себя теоретические вопросы, позволяющие оценить уровень полученных знаний и практические задания, позволяющие оценить степень сформированности умений и навыков.

Для оценивания результатов обучения на экзамене используются следующие показатели:

- 1) знание учебного материала и владение понятийным аппаратом теории машинного обучения;
- 2) умение анализировать многомерные данные и преодолевать вычислительные проблемы, связанные с высокой размерностью данных;
- 3) умение применять методы машинного обучения при решении задач в различных прикладных областях; ;
- 5) владение навыками использования библиотек языка Python для построения систем, обучающихся по прецедентам
- 6) владение навыками построения и проверки качества моделей машинного обучения;
- 7) владение навыками интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов.

Практическое задание

Ответьте на вопросы о данных по авиарейсам в США за январь-апрель 2008 года.

По ссылке расположены [Данные](#) и их [описание](#)

- 1) Читайте выборку из файла при помощи функции `pd.read_csv` и ответьте на следующие вопросы:
 - Имеются ли в данных пропущенные значения?
 - Сколько всего пропущенных элементов в таблице "объект-признак"?
 - Сколько объектов имеют хотя бы один пропуск?
 - Сколько признаков имеют хотя бы одно пропущенное значение?
- 2) Преобразуйте каждый признак `FeatureName` из указанных в пару новых признаков `FeatureName_Hour`, `FeatureName_Minute`, разделив каждое из значений на часы и минуты. Не забудьте при этом исключить исходный признак из выборки. В случае, если значение признака отсутствует, значения двух новых признаков, его заменяющих, также должны отсутствовать.
- 3) Некоторые из признаков, отличных от целевой переменной, могут оказывать чересчур значимое влияние на прогноз, поскольку по своему смыслу содержат большую долю информации о значении целевой переменной. Изучите описание датасета и исключите

признаки, сильно коррелирующие с ответами. Ваш выбор признаков для исключения из выборки обоснуйте.

4) Приведите данные к виду, пригодному для обучения линейных моделей. Для этого вещественные признаки надо отмасштабировать, а категориальные — привести к числовому виду. Также надо устранить пропуски в данных. Реализуйте функцию `transform_data`, которая принимает на вход `DataFrame` с признаками и выполняет следующие шаги:

- Замена пропущенных значений на нули для вещественных признаков и на строки 'nan' для категориальных.
- Масштабирование вещественных признаков с помощью [StandardScaler](#).
- One-hot-кодирование категориальных признаков с помощью [DictVectorizer](#) или функции [pd.get_dummies](#).

Метод должен возвращать преобразованный `DataFrame`, который должна состоять из масштабированных вещественных признаков и закодированных категориальных (исходные признаки должны быть исключены из выборки).

5) Разбейте выборку и вектор целевой переменной на обучение и контроль в отношении 70/30 (для этого можно использовать функцию [train_test_split](#)).

Перечень вопросов к зачету

1. Классификация задач Data Mining. Сферы применения Data Mining
2. Задача классификации. Процесс классификации. Методы, применяемые для решения задач классификации. Точность классификации: оценка уровня ошибок
3. Задача кластеризации Оценка качества кластеризации Процесс кластеризации 4. Применение кластерного анализа Кластерный анализ в маркетинговых исследованиях Практика применения кластерного анализа в маркетинговых исследованиях
5. Задача прогнозирования Сравнение задач прогнозирования и классификации 6. 6. Прогнозирование и временные ряды Тренд, сезонность и цикл
7. Точность прогноза Виды прогнозов Методы прогнозирования
8. Деревья решений.
9. Метод опорных векторов.
10. Метод "ближайшего соседа".
11. Байесовская классификация.
12. Классификация с помощью Нейронных сетей
13. Иерархические методы кластеризации.
14. Итеративные методы кластеризации.
15. Ассоциативные правила. Часто встречающиеся шаблоны или образцы. Поддержка. Характеристики ассоциативных правил. Границы поддержки и достоверности ассоциативного правила.
16. Методы поиска ассоциативных правил Разновидности алгоритма Apriori .
17. Методы визуализации Представление данных в одном, двух и трех измерениях

Для оценивания результатов обучения на зачете с оценкой используется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Соотношение показателей, критериев и шкалы оценивания результатов обучения.

Критерии оценивания	Шкала оценок
<i>Обучающийся в полной мере владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), сдал все практические и лабораторные работы, среднее количество правильных ответов на вопросы тестов превышает 80%.</i>	<i>Отлично</i>
<i>Обучающийся владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), но не сдал одну практическую или лабораторную работу, среднее количество правильных ответов на вопросы тестов находится в диапазоне 70-80%.</i>	<i>Хорошо</i>

<i>Обучающийся демонстрирует неуверенное владение понятийным аппаратом данной области науки (теоретическими основами дисциплины), не сдал две практических или лабораторных работы, среднее количество правильных ответов на вопросы тестов находится в диапазоне 60-70%.</i>	<i>Удовлетворительно</i>
<i>Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не сдал более двух практических или лабораторных работы, среднее количество правильных ответов на вопросы тестов менее 70%.</i>	<i>Неудовлетворительно</i>