

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

УТВЕРЖДАЮ
Заведующий кафедрой

Математических методов исследования операций



Азарнова Т.В.

29 мая 2023 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.О.12 Интеллектуальный анализ данных

1. Код и наименование направления подготовки/специальности:

38.04.05 Бизнес-информатика

2. Профиль подготовки / специализация/магистерская программа:

Информационная бизнес-аналитика

3. Квалификация (степень) выпускника: магистр

4. Форма образования: заочная

5. Кафедра, отвечающая за реализацию дисциплины:

Математических методов исследования операций

6. Составители программы: Замятин Игорь Викторович, к. ф.-м. наук

7. Рекомендована:

Научно-методическим советом факультета прикладной математики, информатики и механики от 26.05.2023 Протокол №7

8. Учебный год: 2023/2024

Семестр(-ы): 1-2

9. Цели и задачи учебной дисциплины:

Целью курса является формирование у студентов представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining) и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности.

Задачи курса:

1. Приобретение знаний по методам и моделям Data Mining;
2. Изучение методов, основанных на нечеткой логике;
3. Получение представления о преимуществах и недостатках нейросетевых технологий;
4. Получение представления об алгоритмах построения деревьев решений;
5. Освоение методов кластеризации;
6. Получение практических навыков в работе с существующими программными пакетами по интеллектуальному анализу данных.

10. Место учебной дисциплины в структуре ООП:

Дисциплина относится к обязательным дисциплинам блока Б1 дисциплин учебного плана. Для освоения дисциплины необходимы знания, умения и навыки, сформированные в ходе освоения дисциплины «Теория вероятностей и математическая статистика». В результате освоения дисциплины формируются знания, умения и навыки, необходимые для изучения дисциплины «Анализ больших данных».

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-3	Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального	ОПК-3.1	Применяет современные методы и программный инструментарий сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта в профессиональной деятельности в решении профессиональных задач	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. Владеть: - навыками интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов.

оборудования и систем искусственного интеллекта	ОПК-3.2	Разрабатывает методы принятия решений и стратегического планирования и прогнозирования на основе современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	<p>Знать:</p> <ul style="list-style-type: none"> - методы предварительной обработки данных (кодирование, стандартизация и нормализация, устранение выбросов, заполнение пропусков); - методы отбора информативных признаков. <p>Уметь:</p> <ul style="list-style-type: none"> - анализировать многомерные данные и преодолевать вычислительные проблемы, связанные с высокой размерностью данных.
	ОПК-3.3	Разрабатывает системы поддержки принятия решений и стратегического планирования и прогнозирования на основе современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	<p>Знать:</p> <ul style="list-style-type: none"> - методы классификации; - методы регрессионного анализа методы анализа текстовых данных. <p>Уметь:</p> <ul style="list-style-type: none"> - использовать библиотеки языка Python для построения моделей машинного обучения. <p>Владеть:</p> <ul style="list-style-type: none"> - навыками построения и проверки качества моделей машинного обучения; - навыками использования библиотек языка Python для построения систем, обучающихся по прецедентам.

12. Объем дисциплины в зачетных единицах/часах в соответствии с учебным планом — 4/144

Форма промежуточной аттестации – экзамен, курсовая работа.

13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость (часы)			
		Всего	В том числе в интерактивной форме	По семестрам	
				1 сем.	2 сем.
Аудиторные занятия		14	8	12	2
в том числе:	лекции	6		6	
	практические				
	лабораторные	8	8	6	2
Самостоятельная работа		121		62	59
Итого:		135	8	74	61
Форма промежуточной аттестации (экзамен – 6 ч., курс. работа – 3 ч.)		9			9

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 Data Mining.	Классификация задач Data Mining. От данных к решениям. Сопоставление и сравнение понятий "информация", "данные", "знание". Сферы применения Data Mining	Б1.О.12 Интеллектуальный анализ данных
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.	Б1.О.12 Интеллектуальный анализ данных
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	Б1.О.12 Интеллектуальный анализ данных
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.	Б1.О.12 Интеллектуальный анализ данных
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)	Б1.О.12 Интеллектуальный анализ данных
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 .	Б1.О.12 Интеллектуальный анализ данных
1.7	Способы визуального представления данных. Методы визуализации.	Визуализация инструментов Data Mining Визуализация Data Mining моделей. Методы визуализации. Качество визуализации.	Б1.О.12 Интеллектуальный анализ данных
3. Лабораторные работы.			
3.1	Классификация и кластеризация.	Анализ практических задач. Определение типа задачи (классификация, регрессия, кластеризация). Подготовка исходных данных.	Б1.О.12 Интеллектуальный анализ данных
3.2	Прогнозирование и визуализация	Разработка модели МО для решения задач регрессии. Предварительный анализ исходных данных. Визуализация признаков. Базовый статистический анализ средствами Python.	Б1.О.12 Интеллектуальный анализ данных
3.3	Методы классификации и прогнозирования. Ансамблирование.	Разработка моделей МО для решения задачи классификации: Деревья решений. Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация. Нейронные сети. Построение ансамблей классификаторов: случайный лес, бустинг, бэггинг.	Б1.О.12 Интеллектуальный анализ данных

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
3.4	Методы кластерного анализа. Отбор значимых признаков.	Разработка моделей МО для решения задачи классификации: K-средних, агломеративная кластеризация, DBSCAN. Расчет критериев качества кластеризации. Использование метода главных компонент для снижения размерности признаков.	Б1.О.12 Интеллектуальный анализ данных
3.5	Методы поиска ассоциативных правил	Разработка моделей МО для решения задачи поиска ассоциативных правил: алгоритм Apriori.	Б1.О.12 Интеллектуальный анализ данных
3.6	Способы визуального представления данных. Методы визуализации.	Визуализация инструментов Data Mining. Методы визуализации.	Б1.О.12 Интеллектуальный анализ данных

13.2 Темы (разделы) дисциплины и виды занятий:

№ п/п	Наименование раздела дисциплины	Виды занятий (часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Задачи Data Mining.	0,5		0	10	10,5
2	Классификация и кластеризация	0,5		1	14	15,5
3	Прогнозирование и визуализация	0,5		1	20	21,5
4	Методы классификации и прогнозирования. Ансамблирование.	2		2	25	29
5	Методы кластерного анализа. Отбор значимых признаков.	1		2	20	23
6	Методы поиска ассоциативных правил	1		1	20	22
7	Способы визуального представления данных. Методы визуализации.	0,5		1	12	13,5
	Итого:	6		8	121	135

14. Методические указания для обучающихся по освоению дисциплины

Работа с конспектами занятий, презентациями, выполнение практических заданий для самостоятельной работы, выполнение лабораторных работ, использование рекомендованной литературы и методических материалов, в том числе размещенных на странице курса «Б1.О.12 Интеллектуальный анализ данных» на портале «Электронный университет ВГУ» <https://edu.vsu.ru/course/view.php?id=6188>, автор Замятин И.В.

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 лабораторных занятиях. К промежуточной аттестации в виде Курсовой работы учащимся предоставляется письменный отчет о выполненном лабораторном исследовании.

При использовании дистанционных образовательных технологий и электронного обучения необходимо 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:

а) основная литература:

№ п/п	Источник
1	Маккинни, У. Python и анализ данных / У. Маккинни; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва: ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст: электронный // Лань: электронно-библиотечная система. — URL: https://e.lanbook.com/book/131721 (дата обращения: 08.12.2020).
2	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва: ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7. — Текст: электронный // Лань: электронно-библиотечная система. — URL: https://e.lanbook.com/book/69955 (дата обращения: 08.12.2020).
3	Замятин И.В. Программирование на языке Python [Электронный ресурс] : учебно-методическое пособие : [для студ. 3-го курса, обучающихся по направлению 38.03.05 - Бизнес-информатика] / И.В. Замятин ; Воронеж. гос. ун-т. — Воронеж: Издательский дом ВГУ, 2019. — Свободный доступ из интрасети ВГУ. — URL: http://www.lib.vsu.ru/elib/texts/method/vsu/m19-160.pdf

б) дополнительная литература:

№ п/п	Источник
4	Козьмо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Козьмо, В. Ричарт; перевод с английского А. А. Слинкин. — 2-е изд. — Москва: ДМК Пресс, 2016. — 302 с. — ISBN 978-5-97060-330-7. — Текст: электронный // Лань: электронно-библиотечная система. — URL: https://e.lanbook.com/book/82818 (дата обращения: 08.12.2020).
5	Каширина И.Л. Нейросетевые технологии : учебно-методическое пособие для вузов / И.Л. Каширина ; Воронеж. гос. ун-т. — Воронеж : ИПЦ ВГУ, 2008. — 70 с. : ил. — <URL: http://www.lib.vsu.ru/elib/texts/method/vsu/m08-110.pdf >.
6	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.
7	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
8	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

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет):

№ п/п	Источник
9	Электронная библиотечная система «Издательства «Лань». Режим доступа: http://e.lanbook.com/
10	Интеллектуальный анализ данных / И.В. Замятин — Образовательный портал «Электронный университет ВГУ». — Режим доступа: https://edu.vsu.ru/course/view.php?id=6188
11	Практикум по машинному обучению / И.В. Замятин — Образовательный портал «Электронный университет ВГУ». — Режим доступа: https://edu.vsu.ru/course/view.php?id=10157
12	Электронная библиотечная система ВГУ. Режим доступа: http://www.lib.vsu.ru

16. Перечень учебно-методического обеспечения для самостоятельной работы

Самостоятельная работа обучающегося должна включать подготовку к лабораторным занятиям, выполнение текущих заданий по освоению соответствующих тем курса, выполнение курсовой работы и подготовку к промежуточной аттестации. Для этого рекомендуется освоить теоретический материал соответствующих тем по конспектам лекций, презентационному материалу, размещенному на ЭО ресурсах, литературу из представленного перечня, материалы с тематических ресурсов сети Интернет.

№ п/п	Источник
-------	----------

1	Интеллектуальный анализ данных / И.В. Замятин — Образовательный портал «Электронный университет ВГУ». — Режим доступа: https://edu.vsu.ru/course/view.php?id=6188
2	Практикум по машинному обучению / И.В. Замятин — Образовательный портал «Электронный университет ВГУ». — Режим доступа: https://edu.vsu.ru/course/view.php?id=10157
3	Электронная библиотечная система ВГУ. Режим доступа: http://www.lib.vsu.ru

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

Python 3 с подключенными библиотеками (дистрибутив Anaconda).

Дисциплина реализуется с применением электронного обучения и дистанционных образовательных технологий. Для организации самостоятельной работы обучающихся используется онлайн-курс «Б1.О.12 Интеллектуальный анализ данных», размещенный на платформе Электронного университета ВГУ (LMS moodle), а также Интернет-ресурсы, приведенные в п.15в.

18. Материально-техническое обеспечение дисциплины:

Лекционная аудитория должна быть оснащена современным компьютером с подключенным к нему проектором на настенный экран. Лабораторные занятия должны проводиться в специализированной аудитории, оснащенной современными персональными компьютерами и программным обеспечением в соответствии с тематикой изучаемого материала.

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 Data Mining.	ОПК-3	ОПК-3.1	Задание для Лабораторной работы №1
2.	Классификация и кластеризация			
3.	Прогнозирование и визуализация		ОПК-3.2	
4.	Методы классификации и прогнозирования. Ансамблирование.		ОПК-3.3	Задание для Лабораторной работы №2
5.	Методы кластерного анализа. Отбор значимых признаков.			Задание для Лабораторной работы №3
6.	Методы поиска ассоциативных правил			
7.	Способы визуального представления данных. Методы визуализации.			Задание для Лабораторной работы №4
Промежуточная аттестация форма контроля – экзамен, курсовая работа				Темы для курсовых работ Список вопросов к экзамену

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:

Лабораторная работа № 1.

- 1) Загрузить один из предложенных преподавателем датасетов. Выполнить предварительную подготовку данных, используя соответствующие методы библиотек Numpy и Pandas.
- 2) Решить поставленную задачу с помощью регрессионных моделей. Осуществить настройку гиперпараметров.
- 3) Проанализировать качество получившихся моделей и выбрать наилучшую, объяснив свой выбор.

Лабораторная работа № 2.

- 1) Загрузить один из предложенных преподавателем датасетов. Выполнить предварительную подготовку данных, используя соответствующие методы библиотек Numpy и Pandas.
- 2) Решить поставленную задачу с помощью моделей классификации. Осуществить настройку гиперпараметров.
- 3) Проанализировать качество получившихся моделей и выбрать наилучшую, объяснив свой выбор.

Лабораторная работа № 3.

- 1) Загрузить один из предложенных преподавателем датасетов. Выполнить предварительную подготовку данных, используя соответствующие методы библиотек Numpy и Pandas.
- 2) Исходя из описания задачи, определить тип задачи машинного обучения и выбрать необходимые виды моделей для ее решения.
- 3) Решить поставленную задачу с помощью ансамблевых методов. Осуществить настройку гиперпараметров.
- 4) Проанализировать качество получившихся моделей и выбрать наилучшую, объяснив свой выбор.

Лабораторная работа № 4.

- 1) Загрузить предложенный преподавателем датасет. Выполнить предварительную подготовку данных, используя соответствующие методы библиотек Numpy и Pandas.
- 2) Построить графики, отражающий основные статистические свойства исходной выборки. Использовать различные виды графиков (точечные, линейные, столбчатые, круговые диаграммы).
- 3) Вывести основные описательные статистики (средняя, медиана, корреляции и т.п.).

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: Контрольная работа, Экзамен. 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.

Промежуточная аттестация проводится в соответствии с Положением о промежуточной аттестации обучающихся по программам высшего образования.

Перечень вопросов к экзамену:

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.
18. Параллельные координаты. "Лица Чернова". Качество визуализации. Представление пространственных характеристик.

Оценка результатов обучения на **Экзамене** может производиться также с помощью тестирования:

Вопросы с выбором ответа

1. С помощью метода деревьев решений возможно решение задач:
 - a. Регрессии.
 - b. Классификации.
 - c. Классификации и регрессии.**
 - d. Кластеризации.

2. Какой из видов задач машинного обучения направлен на предсказание значения той или иной непрерывной числовой величины на основе входных данных?
 - a. Регрессия.**
 - b. Классификация.
 - c. Кластеризация.
 - d. Переобучение.

3. Обучение без учителя предполагает отсутствие...

a. правильных выходов модели.

b. обучающего множества.

c. эксперта, корректирующего процесс обучения.

d. алгоритма обучения.

4. Множество примеров, используемое для обучения модели, называется...

a. обучающим множеством

b. валидационным множеством

c. тестовым множеством

5. Множество примеров, используемое для проверки работы сконструированной модели, называется...

a. тестовой выборкой

b. синтетической выборкой

c. обучающей выборкой

6. Обучающая выборка в задаче обучения с учителем — это...

a. набор данных, каждая запись которого представляет собой учебный пример, содержащий заданное входное влияние, и соответствующий ему правильный выходной результат

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

c. группировка объектов на основе данных, описывающих свойства объектов

7. Что такое стандартизация данных?

a. Перевод значений всех признаков в числовую шкалу измерения

b. Выделение значимых признаков

c. Преобразование входных признаков, при котором все входные значения масштабируются в отрезок $[0,1]$

d. Преобразование входных признаков так, чтобы среднее значение каждого признака было 0, а стандартное отклонение 1

8. Задача классификации сводится к ...

- a. поиску независимых групп и их характеристик в всем множестве анализируемых данных
- b. определению по известным характеристикам объекта значение некоторого его параметра
- c. нахождению частых зависимостей между объектами или событиями

d. определению класса объекта по его характеристикам

9. Что такое бэггинг?

- a. метод кодирования нечисловых данных
- b. процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих
- c. один из методов регуляризации в задачах машинного обучения

d. процедура независимого построения композиции алгоритмов машинного обучения

10. Что такое бустинг?

- a. метод кодирования нечисловых данных
- b. процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов**
- c. процедура независимого построения композиции алгоритмов машинного обучения
- d. метод борьбы с переобучением в задачах машинного обучения

11. Что из перечисленного является признаком переобучения модели?

- a. когда ошибка на обучающей выборке низкая, а на тестовой высокая**
- b. когда ошибка на обучающей выборке высокая, а на тестовой низкая
- c. когда ошибка на обучающей и на тестовой выборке высокая
- d. когда ошибка на обучающей и на тестовой выборке низкая

12. Переобучение - это ...

- a. излишнее обучение модели, не дающее прироста точности
- b. повторное обучение модели для проверки ее корректности
- c. дообучение модели на новых данных
- d. когда построенная модель хорошо объясняет примеры из обучающей выборки, но плохо работает на новых данных**

13. При решении задачи регрессии ищут...

- a. правила или набор правил в соответствии с которыми любой новый объект можно отнести к одному из классов
- b. функциональные зависимости, которые позволяют прогнозировать изменения непрерывных числовых параметров**
- c. группы, на которые можно разделить объекты, данные о которых подвергаются анализу
- d. соотношения между зависимыми и независимыми показателями и переменными в наглядной и понятной человеку форме

14. С помощью каких методов определяют коэффициенты уравнения регрессии:

- a. метода наименьших модулей
- b. метода Гаусса
- c. метода наименьших квадратов**
- d. симплекс-метода

15. Что такое перекрестная проверка (Cross-validation)?

(укажите все правильные варианты):

- a. Метод формирования обучающего и тестового множеств для обучения модели в условиях недостаточности исходных данных или неравномерного представления классов.**
- b. Построение нескольких моделей для одного исходного набора данных.
- c. Метод оценки эффективности выбранной модели с наиболее равномерным использованием имеющихся данных.**

16. Явление переобучения характеризуется ...

(укажите все правильные варианты):

- a. слишком долгим обучением
- b. чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению**
- c. существенным ростом весовых коэффициентов модели

Открытые вопросы

17. Чему равно значение ошибки MSE для выборки из двух обучающих векторов, если реальные выходы модели $Z=(z_1,z_2)=(0.27, 0.43)$, а целевые (правильные) выходы $Y=(y_1,y_2)=(0.29, 0.23)$?

- a. 0,0202
- b. Критерий оценивания: точное соответствие

18. Дана матрица ошибок, построенная по результатам работы некоторого алгоритма классификации. Общая точность (ассигасу) равна...

		Истинный класс	
		1	-1
Предсказанный класс	1	25	20
	-1	20	15

a. 0,5

- b. Критерий оценивания: точное соответствие

Критерии оценивания для вопросов №1-14: указан правильный вариант – 1, в противном случае – 0.

Критерии оценивания для вопросов №15, 16: указаны оба правильных варианта – 1, указан один из правильных вариантов – 0,5, не указано ни одного правильного варианта – 0.

Задание для Курсовой работы:

- 1) Загрузить один из предложенных преподавателем датасетов. Выполнить предварительную подготовку данных, используя соответствующие методы библиотек Numpy и Pandas.
- 2) Исходя из описания задачи, определить тип задачи машинного обучения и выбрать необходимые виды моделей для ее решения.
- 3) Построить модели машинного обучения для решения поставленной задачи. Осуществить настройку гиперпараметров.
- 4) Проанализировать качество получившихся моделей и выбрать наилучшую, объяснив свой выбор.
- 5) Составить письменный Отчет о выполнении Курсовой работы.

Для оценивания результатов обучения на **Экзамене** используются следующие показатели:

- 1) знание основных задач интеллектуального анализа данных и владение понятийным аппаратом;
- 2) знание моделей и методов интеллектуального анализа данных;

- 3) умение применять теоретические знания для решения практических задач;
- 4) умение иллюстрировать ответ примерами, фактами, данными научных исследований.

Для оценивания результатов обучения в **Курсовой работе** используются следующие показатели:

- 1) умение применять модели и методы машинного обучения для решения практических задач интеллектуального анализа данных;
- 2) владение технологиями разработки алгоритмов и программными системами интеллектуального анализа данных.

Для оценивания результатов обучения на Экзамене и по Курсовой работе используется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Соотношение показателей, критериев и шкалы оценивания результатов обучения на **Экзамене**:

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Обучающийся в полной мере владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), способен иллюстрировать ответ примерами, фактами, данными научных исследований, применять теоретические знания для решения практических задач в области интеллектуального анализа данных. По итогам решения Самостоятельных работ Обучающимся получены хорошие или отличные оценки.	<i>Повышенный уровень</i>	<i>Отлично</i>
Ответ на контрольно-измерительный материал не соответствует одному из перечисленных показателей, но обучающийся дает правильные ответы на дополнительные вопросы. По итогам решения Самостоятельных работ Обучающимся получены хорошие или отличные оценки.	<i>Базовый уровень</i>	<i>Хорошо</i>
Ответ на контрольно-измерительный материал не соответствует любым двум из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. По итогам решения Самостоятельных работ Обучающимся получены хорошие или удовлетворительные оценки.	<i>Пороговый уровень</i>	<i>Удовлетворительно</i>
Ответ на контрольно-измерительный материал не соответствует любым трем из перечисленных показателей. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки. По итогам решения Самостоятельных работ Обучающимся получены оценки не выше удовлетворительных.	–	<i>Неудовлетворительно</i>

Соотношение показателей, критериев и шкалы оценивания результатов обучения по **Курсовой работе**:

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
---------------------------------	--------------------------------------	--------------

	й	
Обучающийся в полной мере владеет технологиями разработки алгоритмов и программными системами интеллектуального анализа данных, умеет применять модели и методы машинного обучения для решения практических задач, способен качественно представить (оформить) результат самостоятельной научно-практической работы.	<i>Повышенный уровень</i>	<i>Отлично</i>
Обучающийся в полной мере владеет технологиями разработки алгоритмов и программными системами интеллектуального анализа данных, умеет применять модели и методы машинного обучения для решения практических задач. Однако Отчет о Курсовой работе оформлен некачественно, тема раскрыта не полностью, есть замечания по структуре изложения.	<i>Базовый уровень</i>	<i>Хорошо</i>
Ответ на контрольно-измерительный материал не соответствует одному из перечисленных показателей, но Отчет о Курсовой работе оформлен качественно, тема раскрыта, нет замечаний по структуре изложения.	<i>Пороговый уровень</i>	<i>Удовлетворительно</i>
Ответ на контрольно-измерительный материал не соответствует одному из перечисленных показателей. Отчет о Курсовой работе оформлен некачественно, тема не раскрыта, есть замечания по структуре изложения.	–	<i>Неудовлетворительно</i>

Задания раздела 20.2 рекомендуются к использованию при проведении диагностических работ с целью оценки остаточных результатов освоения данной дисциплины (знаний, умений, навыков).