

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.
26.05.2020



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Б1.В.02 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: НМС факультета Прикладной математики, информатики и механики, протокол №9 от 23.05.2020
8. Учебный год: 2020/2021 Семестр(-ы): 1

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 вариативной части 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 выпускников):

Компетенция		Планируемые результаты обучения
Код	Название	
ОК-1	способность к абстрактному мышлению, анализу, синтезу	знать: основные задачи и методы интеллектуального анализа данных уметь: формулировать задачи интеллектуального анализа данных, выбирать адекватные алгоритмы их решения, оценивать качество получаемых решений владеть (иметь навык(и)): технологиями разработки алгоритмов и программными системами интеллектуального анализа данных

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

Форма промежуточной аттестации – экзамен, курсовая работа.

13. Виды учебной работы:

Вид учебной работы	Трудоемкость (часы)		
	Всего	В том числе в интерактивной форме	По семестрам
1 сем.			
Аудиторные занятия	16	8	16
в том числе:	лекции	8	8
	практические		
	лабораторные	8	8

Самостоятельная работа	119		119
Итого:	135	8	135
Форма промежуточной аттестации (экзамен – 6 ч., курс. работа – 3 ч.)	9		9

13.1 Содержание дисциплины:

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
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
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.
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
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.
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)
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 .
1.7	Способы визуального представления данных. Методы визуализации.	Визуализация инструментов Data Mining Визуализация Data Mining моделей. Методы визуализации. Качество визуализации.
3. Лабораторные работы.		
3.1	Классификация и кластеризация.	Анализ практических задач. Определение типа задачи (классификация, регрессия, кластеризация). Подготовка исходных данных.
3.2	Прогнозирование и визуализация	Разработка модели МО для решения задач регрессии. Предварительный анализ исходных данных. Визуализация признаков. Базовый статистический анализ средствами Python.
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.

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
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.
3.5	Методы поиска ассоциативных правил	Разработка моделей МО для решения задачи поиска ассоциативных правил: алгоритм Apriori.
3.6	Способы визуального представления данных. Методы визуализации.	Визуализация инструментов Data Mining. Методы визуализации.

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.	1		0	10	11
2	Классификация и кластеризация	1		1	14	16
3	Прогнозирование и визуализация	1		1	20	22
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.	1		1	10	12
	Итого:	8		8	119	135

14. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины «Интеллектуальный анализ данных» включает лекционные и лабораторные занятия, самостоятельную работу обучающихся. Рекомендуется конспектировать лекционный материал, смотреть презентации по соответствующей теме, чтобы систематизировать изучаемый материал. Необходимо активное самостоятельное изучение методической и учебной литературы по предлагаемому перечню. В течение семестра обучающимся предлагается выполнить самостоятельные задания. К промежуточной аттестации в виде Курсовой работы учащимися 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	Коэльо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Коэльо, 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)

Дисциплина реализуется с применением электронного обучения и дистанционных образовательных технологий. Для организации самостоятельной работы обучающихся используется онлайн-курс «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. Фонд оценочных средств:

19.1 Перечень компетенций с указанием этапов формирования и планируемых результатов обучения

Код и содержание компетенции (или ее части)	Планируемые результаты обучения (показатели достижения заданного уровня освоения компетенции посредством формирования знаний, умений, навыков)	Этапы формирования компетенции (разделы (темы) дисциплины или модуля и их наименование)	ФОС* (средства оценивания)
ОК-1 способность к абстрактному мышлению, анализу, синтезу	Знать: основные задачи и методы интеллектуального анализа данных	1.1 Задачи Data Mining. 1.2 Классификация и кластеризация. 1.3 Прогнозирование и визуализация. 1.4 Методы классификации и прогнозирования. Ансамблирование. 1.5 Методы кластерного анализа. Отбор значимых признаков. 1.6 Методы поиска ассоциативных правил.	Список вопросов к Экзамену. Задание для Курсовой работы
	Уметь: формулировать задачи интеллектуального анализа данных, выбирать	1.7 Способы визуального представления данных. Методы визуализации. 3.6 Способы визуального представления данных. Методы визуализации.	Задание для Самостоятельной работы № 1

	адекватные алгоритмы их решения, оценивать качество получаемых решений	3.1 Классификация и кластеризация. 3.2 Прогнозирование и визуализация. 3.3 Методы классификации и прогнозирования. Ансамблирование. 3.4 Методы кластерного анализа. Отбор значимых признаков. 3.5 Методы поиска ассоциативных правил.	Задание для Самостоятельной работы № 2
Промежуточная аттестация			КИМ

19.2 Описание критериев и шкалы оценивания компетенций (результатов обучения) при промежуточной аттестации

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	<i>Базовый уровень</i>	<i>Хорошо</i>

вопросы. По итогам решения Самостоятельных работ Обучающимся получены хорошие или отличные оценки.		
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
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>

19.3 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения образовательной программы

19.3.1 Перечень вопросов к экзамену (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. 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.

19.3.2 Перечень заданий для контрольных работ

Самостоятельная работа № 1.

- 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.).

Самостоятельная работа № 2.

- 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.

19.3.3 Темы курсовых работ

- 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.

19.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Оценка знаний, умений и навыков, характеризующая этапы формирования компетенций в рамках изучения дисциплины осуществляется в ходе текущей и промежуточной аттестаций.

Текущая аттестация проводится в соответствии с Положением о текущей аттестации обучающихся по программам высшего образования Воронежского государственного университета. Текущая аттестация проводится в форме письменных работ (самостоятельные лабораторные работы), а также оценки результатов практической деятельности (курсовая работа). Критерии оценивания приведены выше.

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

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

При оценивании используются количественные шкалы оценок. Критерии оценивания приведены выше.