

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

УТВЕРЖДАЮ
Заведующий кафедрой
математического моделирования



М.Ш. Бурлуцкая

26.06.2022 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.О.02.14 Машинное обучение

1. Код и наименование специальности:

10.05.04 Информационно-аналитические системы безопасности

2. Специализация:

Автоматизация информационно-аналитической деятельности

3. Квалификация выпускника: Специалист по защите информации

4. Форма обучения: Очная

5. Кафедра, отвечающая за реализацию дисциплины:

Кафедра математического моделирования

6. Составитель программы: Костин Дмитрий Владимирович, д.ф.-м.н.

7. Рекомендована: Научно-методическим советом математического факультета, протокол № 0500-03 от 24.03.2022

8. Учебный год: 2027/2028

Семестр: В (11)

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются:

- ознакомление студентов с теоретическими основами и принципами машинного обучения;
- формирование теоретических знаний в области разработки интеллектуальных информационных систем, использующих аппарат машинного обучения, которые позволяют решать практические задачи анализа данных в исследованиях и бизнес-приложениях.

Задачи учебной дисциплины:

- формирование у студентов практических навыков работы с данными и решения прикладных задач анализа больших данных;
- изучение моделей и классических алгоритмов машинного обучения (линейные, метрические, логические и байесовские модели) и формирование навыков их использования в профессиональной деятельности;
- приобретение навыков в области оценки адекватности построенных моделей.

10. Место учебной дисциплины в структуре ОПОП:

Дисциплина «Машинное обучение» относится к обязательной части Блока 1. Дисциплины (модули).

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-10	Способен разрабатывать и применять математические модели и методы анализа массивов данных и интерпретировать профессиональный смысл получаемых формальных результатов	ОПК-10.4	Способен применять методы машинного обучения для решения задач распознавания, классификации, прогнозирования	Знать: основные понятия машинного обучения; Уметь: применять различные методы машинного обучения для создания программ анализа больших массивов данных; Владеть: навыками решения практических задач с использованием методов машинного обучения.

12. Объем дисциплины в зачетных единицах/час. — 4/144.

Форма промежуточной аттестации: зачет с оценкой.

13. Трудоемкость по видам учебной работы

Вид учебной работы	Трудоемкость		
	Всего	По семестрам	
		Семестр В (11)	
Контактная работа	90	90	
в том числе:	лекции	50	50
	практические	0	0

	лабораторные	40	40	
	курсовая работа			
	контрольные работы			
Самостоятельная работа		54	54	
Промежуточная аттестация				
Итого:		144	144	

13.1. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
1. Лекции			
1.1	Типы задач. Метрические классификаторы. Алгоритмы кластеризации	Типы задач машинного обучения. Предмет и задачи машинного обучения и анализа данных. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Основные этапы эволюции алгоритмов машинного обучения. Метрические классификаторы. Общий вид метрического классификатора. Алгоритм К ближайших соседей. Алгоритмы отбора эталонов. Алгоритмы кластеризации. Алгоритмы кластеризации с фиксированным количеством кластеров. Алгоритмы кластеризации по плотности. Иерархическая кластеризация.	
1.2	Деревья решений, линейные классификаторы. Нейронные сети	Деревья решений. Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев. Линейные классификаторы. Перцептрон и разделяющая гиперплоскость. Переход в пространство повышенной размерности. Метод опорных векторов. Нейронные сети и глубокое обучение. Логистическая регрессия. Градиентный спуск. Нейронные сети и алгоритм обратного распространения градиента. Глубокое обучение, свертки и пулинг.	
1.3	Регрессионный анализ, Ансамблевые методы. Стохастический поиск	Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия. Гребневая регрессия. Ансамблевые методы. Голосование. Бутстраппинг. Бустинг, адаптивный бустинг, градиентный бустинг. Стохастический поиск. Монте-Карло поиск. Алгоритм симулированного отжига. Генетический алгоритм.	
2. Лабораторные занятия			
2.1	Обучение нейрона логическим функциям	Реализация алгоритма обучения одного нейрона функциям конъюнкции, дизъюнкции, отрицания	
2.2	Обучение двуслойной сети логической функции XOR	Реализация алгоритма обучения двуслойного перцептрона функции «исключающие или»	
2.3	Обучение методом градиентного спуска	Реализация алгоритма обучения глубокой нейронной сети с использованием методов градиентного спуска	
2.4	Линейная регрессия	Вычисление значения функции Розенброка с использованием реализации алгоритма линейной регрессии на основе нейронной сети	

2.5	Задача классификации. Ирисы Фишера	Реализация алгоритма обучения нейронной сети классификации Ирисов Фишера по заданному набору данных.	
-----	------------------------------------	--	--

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Типы задач	2	-	2	4	8
2	Метрические классификаторы	4	-	4	4	12
3	Алгоритмы кластеризации	6	-	4	6	16
4	Деревья решений	8	-	6	8	22
5	Линейные классификаторы	4	-	4	4	12
6	Нейронные сети	10	-	8	12	30
7	Регрессионный анализ	6	-	4	6	16
8	Ансамблевые методы	6	-	4	6	16
9	Стохастический поиск	4	-	4	4	12
	Итого:	50	-	40	54	144

14. Методические указания для обучающихся по освоению дисциплины:

Освоение дисциплины предполагает не только обязательное посещение обучающимся аудиторных занятий (лекций и лабораторных занятий) и активную работу на них, но и самостоятельную учебную деятельность, на которую отводится 54 часа.

Самостоятельная учебная деятельность студентов по дисциплине «Машинное обучение» предполагает выполнение следующих заданий:

1) самостоятельное изучение учебных материалов по разделам дисциплины с использованием основной и дополнительной литературы, информационно-справочных и поисковых систем;

2) подготовку к текущим аттестациям: выполнение лабораторных заданий по поиску необходимых для работы в аудитории материалов в Интернете.

Особое внимание обучающихся направляется на освоение методов машинного обучения, построение алгоритмов и решения практических задач анализа больших данных.

Все выполняемые студентами самостоятельно задания подлежат последующей проверке преподавателем для получения допуска к зачету.

В случае необходимости перехода на дистанционный режим обучения будет создан электронный курс «Название» на портале «Электронный университет ВГУ»: : <https://edu.vsu.ru/> . Там же будут размещены необходимые для усвоения курса материалы.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Коэльо Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Коэльо, В. Ричарт ; пер. с англ. Слинкин А.А.. — Электрон. дан. — Москва : ДМК Пресс, 2016. — 302 с. — Режим доступа: https://e.lanbook.com/book/82818 .
2	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 .

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 .
---	---

б) дополнительная литература:

№ п/п	Источник
4	Бринк Х. Машинное обучение / Х. Бринк, Д. Ричардс, М. Феверолф. - СПб.: Питер, 2017. -336 с.
5	Вьюгин В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс] : учебное пособие / В.В. Вьюгин. — Электрон. дан. — Москва : МЦНМО, 2013. — 304 с. — Режим доступа: https://e.lanbook.com/book/56397 .
6	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас. — СПб.: Питер, 2018. — 576 с.
7	Силен Дэви. Основы Data Science и Big Data. Python и наука о данных / Дэви Силен, Арно Мейсман, Мохамед Али. -СПб.: Питер, 2017. -336 с.
8	Шарден Б. Крупномасштабное машинное обучение вместе с Python [Электронный ресурс] : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; пер. с англ. А. В. Логунова. — Электрон. дан. — Москва : ДМК Пресс, 2018. — 358 с. — Режим доступа: https://e.lanbook.com/book/10583 .

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

№ п/п	Ресурс
9	Открытый курс машинного обучения https://habr.com/company/ods/blog/322626/
10	Электронный каталог ЗНБ ВГУ : http://www.lib.vsu.ru .
11	https://math.vsu.ru/wp/?page_id=937 – раздел на сайте математического факультета, на котором размещены методические издания.
12	ЭБС «Университетская библиотека онлайн».
13	Электронный университет ВГУ : https://edu.vsu.ru/ .

16. Перечень учебно-методического обеспечения для самостоятельной работы:

№ п/п	Источник
1	UCI Machine Learning Repository — репозиторий наборов данных для выполнения лабораторных работ по курсу машинного обучения - http://archive.ics.uci.edu/ml/ .
2	Бринк Х. Машинное обучение / Х. Бринк, Д. Ричардс, М. Феверолф. - СПб.: Питер, 2017. -336 с.: Материалы к книге: https://github.com/brinkar/real-world-machine-learning .
3	Курс «Машинное обучение» на портале «Электронный университет ВГУ», автор Каширина И.Л. - https://edu.vsu.ru/course/view.php?id=3579
4	Мюллер А. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо. – 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 .
5	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас. — СПб.: Питер, 2018. — 576 с. - Материалы к книге: https://github.com/jakevdp/PythonDataScienceHandbook .

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ, электронное обучение (ЭО), смешанное обучение):

Дисциплина может реализовываться с применением дистанционных образовательных технологий, например, на платформе «Электронный университет ВГУ» (<https://edu.vsu.ru>).

Перечень необходимого программного обеспечения: Win10pro или Linux, Microsoft Office, LibreOffice 6, Calc, Microsoft Visual Studio, Microsoft Visual C++, Foxit Reader, браузер Mozilla Firefox, Opera или Internet.

18. Материально-техническое обеспечение дисциплины:

Учебная аудитория для проведения занятий лекционного и семинарского типа, текущего контроля и промежуточной аттестации со специализированной мебелью.

Для самостоятельной работы используется класс с компьютерной техникой, оснащенный необходимым программным обеспечением, электронными учебными пособиями и законодательно-правовой и нормативной поисковой системой, имеющий выход в глобальную сеть.

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	Типы задач	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 1, 2, 3, 4, 5
2	Метрические классификаторы	ОПК-10	ОПК-10.4	Комплект лабораторных заданий №5
3	Алгоритмы кластеризации	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 3
4	Деревья решений	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 3, 4, 5
5	Линейные классификаторы	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 5
6	Нейронные сети	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 1, 2, 3, 4, 5
7	Регрессионный анализ	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 4
8	Ансамблевые методы	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 1, 2, 3
9	Стохастический поиск	ОПК-10	ОПК-10.4	Комплект лабораторных заданий № 3, 4, 5
Промежуточная аттестация Форма контроля – зачет с оценкой				Перечень вопросов к дифференцированному зачету

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: устных опросов, проверки домашних заданий, лабораторных работ.

Перечень лабораторных работ:

- Л.р. №1 Обучение нейрона логическим функциям;
- Л.р. №2 Обучение двуслойной сети логической функции XOR;
- Л.р. №3 Обучение методом градиентного спуска;
- Л.р. №4 Линейная регрессия;
- Л.р. №5 Задача классификации. Ирисы Фишера.

Для оценивания текущего контроля успеваемости используются следующие **показатели**:

- 1) знание основных понятий и методов;
- 2) умение применять полученные знания и навыки для решения задач, проводить анализ полученных решений;
- 3) владение математическим аппаратом и современными методами в теории машинного обучения;
- 4) умение писать код компьютерных программ для решения типовых математических задач;
- 5) знание имеющихся ресурсов для решения прикладных математических задач;

- 6) умение использовать стандартные пакеты программного обеспечения для решения типовых математических задач;
- 7) владение навыками хранения, поиска, сбора, систематизации, обработки и использования информации.

Шкала оценок:

Зачтено: Выполнение заданий соответствует перечисленным показателям, обучающийся дает ответы на дополнительные вопросы, может быть не совсем полные. Демонстрирует умение решать задачи, возможно с некоторыми ошибками.

Не зачтено: Ответы не соответствуют ни одному из перечисленных показателей. Обучающийся демонстрирует фрагментарные знания и умения или их отсутствие.

20.2. Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется в форме собеседования по экзаменационным билетам с помощью нижеприведенных оценочных средств (перечень вопросов к зачету).

В билет включаются теоретический вопрос и задача лабораторной работы.

Перечень вопросов к дифференцированному зачету:

№ п/п	Вопросы
1.	Типы задач машинного обучения. Предмет и задачи машинного обучения и анализа данных. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Основные этапы эволюции алгоритмов машинного обучения.
2.	Деревья решений. Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев.
3.	Линейные классификаторы. Перцептрон и разделяющая гиперплоскость. Переход в пространство повышенной размерности. Метод опорных векторов
4.	Нейронные сети и глубокое обучение.
5.	Логистическая регрессия. Градиентный спуск. Нейронные сети и алгоритм обратного распространения градиента.
6.	Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия. Гребневая регрессия.
7.	Ансамблевые методы. Голосование. Бутстраппинг. Бустинг, адаптивный бустинг, градиентный бустинг.
8.	Стохастический поиск. Монте-Карло поиск. Алгоритм симулированного отжига. Генетический алгоритм.

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

- 1) знание теоретических основ;
- 2) умение решать задачи;
- 3) умение работать с алгоритмами методов и информационными ресурсами;
- 4) успешное прохождение текущей аттестации.

Для оценивания результатов на дифференцированном зачете используется **шкала:** «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Соотношение показателей, критериев и шкалы оценивания результатов обучения показаны в следующей таблице:

Критерии оценивания	Шкала оценок
Ответ соответствует всем перечисленным выше показателям, обучающийся дает ответы на дополнительные вопросы. Демонстрирует знание учебного материала.	«Отлично»
Ответ соответствует двум или более из перечисленных показателей, обучающийся дает ответы на дополнительные вопросы, может быть не совсем полные. Демонстрирует знание учебного материала, возможно с некоторыми ошибками.	«Хорошо»
Ответ соответствует одному из перечисленных показателей, обучающийся не дает ответы на дополнительные вопросы. Демонстрирует знание учебного материала с некоторыми ошибками.	«Удовлетворительно»
Ответ не соответствует ни одному из перечисленных показателей. Обучающийся демонстрирует фрагментарные знания и умения или отсутствие их.	«Неудовлетворительно»