

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный университет»

УТВЕРЖДАЮ

Заведующий кафедрой программного обеспечения и администрирования информационных систем



Артемов М. А.

08.06.2018 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ Б1.В.06 Введение в машинное обучение

1. Шифр и наименование направления подготовки:

02.04.03 Математическое обеспечение и администрирование информационных систем

2. Магистерская программа: Информационные технологии

3. Квалификация (степень) выпускника: магистр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: программного обеспечения и администрирования информационных систем

6. Составители программы:

Барановский Евгений Сергеевич, кандидат физико-математических наук

7. Рекомендована: НМС факультета ПММ протокол № 10 от 18.06.2018 г.

8. Учебный год: 2018/2019

Семестр: 1

9. Цели и задачи учебной дисциплины: Формирование знаний, умений и компетенций в области машинного обучения. Изучение различных математических моделей данных и алгоритмов анализа данных. Формирование практических навыков реализации алгоритмов машинного обучения на языке программирования Python с использованием пакетов NumPy, Pandas, Matplotlib и Scikit-Learn.

10. Место учебной дисциплины в структуре ООП: дисциплина относится к вариативной части блока Б1. Для успешного освоения дисциплины необходимо предварительное изучение курса математического анализа, алгебры и дифференциальных уравнений.

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	
ОПК-4	Способность к определению общих форм и закономерностей отдельной предметной области.	знать: особенности различных областей приложения методов машинного обучения; уметь: формализовывать данные для разработки приложений; владеть: технологиями применения методов машинного обучения в различных областях.
ОПК-6	Способность математически корректно ставить естественнонаучные задачи, знать постановки классических задач математики.	знать: математические основы алгоритмов классификации, кластеризации и регрессии; уметь: анализировать применимость алгоритмов машинного обучения к конкретной задаче; владеть: методами разработки моделей конкретной задачи для применения методов машинного обучения.

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

13. Виды учебной работы

Вид учебной работы	Трудоемкость (часы)	
	Всего	По семестрам
		1
Аудиторные занятия	34	34
в том числе:		
лекции		
практические		
лабораторные	34	34
Самостоятельная работа	110	110
Экзамен		
Итого:	144	144

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
		1. Лекции

2. Практические занятия		
3. Лабораторные работы		
3.1	Введение. Оболочка IPython, Jupiter	Введение. Горячие клавиши, магические команды. Ошибки и отладка
3.2	Библиотека NumPy	Типы данных, массивы, вычисления над массивами. Агрегирование, транслирование, сравнение, булева логика. Индексация, сортировка. Структурированные данные.
3.3	Библиотека Pandas	Объекты. Индексация и выборка данных. Операции над данными. Индексация. Объединение. Агрегирование. Векторизованные операции. Временные ряды.
3.4	Библиотека Matplotlib	Графики и диаграммы. Гистограммы. Погрешности. Настройки легенд и шкал цветов. Трехмерные графики.
3.5	Библиотека Scikit-Learn и машинное обучение	Гиперпараметры и проверка модели. Проектирование признаков. Наивная байесовская классификация. Линейная регрессия. Метод опорных векторов. Деревья решения и случайные леса. Метод главных компонент. Метод k-средних. Смеси Гауссовых распределений. Ядерная оценка плотности распределения. Конвейер распознавания лиц.
3.6	Введение. Оболочка IPython, Jupiter	Введение. Горячие клавиши, магические команды. Ошибки и отладка
3.7	Библиотека NumPy	Типы данных, массивы, вычисления над массивами. Агрегирование, транслирование, сравнение, булева логика. Индексация, сортировка. Структурированные данные.
3.8	Библиотека Pandas	Объекты. Индексация и выборка данных. Операции над данными. Индексация. Объединение. Агрегирование. Векторизованные операции. Временные ряды.
3.9	Библиотека Matplotlib	Графики и диаграммы. Гистограммы. Погрешности. Настройки легенд и шкал цветов. Трехмерные графики.
3.10	Библиотека Scikit-Learn и машинное обучение	Гиперпараметры и проверка модели. Проектирование признаков. Наивная байесовская классификация. Линейная регрессия. Метод опорных векторов. Деревья решения и случайные леса. Метод главных компонент. Метод k-средних. Смеси Гауссовых распределений. Ядерная оценка плотности распределения. Конвейер распознавания лиц.

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Введение. Оболочка IPython, Jupiter			8	22	30
2	Библиотека NumPy			8	22	30
3	Библиотека Pandas			8	22	30
4	Библиотека Matplotlib			8	22	30

5	Библиотека Scikit-Learn и машинное обучение			2	22	24
	Итого:			34	110	144

14. Методические указания для обучающихся по освоению дисциплины При изучении дисциплины рекомендуется использовать следующие средства:

- рекомендуемую основную и дополнительную литературу;
- методические указания и пособия;
- контрольные задания для закрепления теоретического материала;
- электронные версии учебников и методических указаний для выполнения практических работ.

Форма организации самостоятельной работы: подготовка к аудиторным занятиям; выполнение домашних заданий; выполнение контрольных работ.

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

а) основная литература:

№ п/п	Источник
1	Златопольский, Д. М. Основы программирования на языке Python / Д. М. Златопольский .— Москва : ДМК Пресс, 2017 .— ISBN 978-5-97060-552-3 .— <URL: https://e.lanbook.com/book/97359 >.
2	Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук .— Москва : ДМК Пресс, 2018 .— ISBN 978-5-97060-508-0 .— <URL: https://e.lanbook.com/book/97353 >.

б) дополнительная литература:

№ п/п	Источник
3	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.
4	Андреас Мюллер, Сара Гвидо Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными - ИД Вильямс, O'Reilly, 2017. – 480 с.

в) информационные электронно-образовательные ресурсы:

№ п/п	Источник
5	www.lib.vsu.ru –ЗНБ ВГУ

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	https://github.com/Gewissta/Introduction_to_ML_with_Python_russian_translation
2	Саммерфилд, М. Python на практике / М. Саммерфилд .— Москва : Лань, 2014 .— 338 с. — ISBN 978-5-97060-095-5 .— <URL: http://e.lanbook.com/book/66480#authors >.

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

18. Материально-техническое обеспечение дисциплины: лекционная аудитория, оснащённая мультимедийным проектором, компьютерный класс с необходимым программным обеспечением.

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)
ОПК-4	Знать: особенности различных областей приложения методов машинного обучения.	Разделы 1-5	Письменный опрос
	Уметь: формализовывать данные для разработки приложений.	Разделы 1-5	Лабораторные работы 1-12
	Владеть: технологиями применения методов машинного обучения в различных областях.	Разделы 1-5	Лабораторные работы 1-12
ОПК-6	Знать: математические основы алгоритмов классификации, кластеризации и регрессии.	Разделы 1-5	Письменный опрос
	Уметь: анализировать применимость алгоритмов машинного обучения к конкретной задаче.	Разделы 1-5	Лабораторные работы 1-12
	Владеть: методами разработки моделей конкретной задачи для применения методов машинного обучения.	Разделы 1-5	Лабораторные работы 1-12
Промежуточная аттестация			По результатам текущих аттестаций

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

Для оценивания результатов обучения на экзамене используются следующие показатели:

- 1) владение основными конструкциями языка Python;
- 2) знание основных возможностей пакета NumPy и умение его использовать в практических задачах;
- 3) знание основных возможностей пакета Pandas и умение его использовать в практических задачах;
- 4) знание основных возможностей пакета Matplotlib и умение его использовать в практических задачах;
- 5) знание основных возможностей пакета Scikit-Learn и умение его использовать в практических задачах;
- 6) владение методами проектирования признаков;
- 7) знание наивных байесовских моделей;
- 8) знание линейных регрессионных моделей;

- 9) знание алгоритма метода опорных векторов;
- 10) владение алгоритмами на основе деревьев решений и случайных лесов;
- 11) знание метода главных компонент;
- 12) умение применять алгоритмы кластеризации методами k-средних;
- 13) умение использовать кластеризацию на основе смеси Гауссовых распределений;
- 14) умение применять методы машинного обучения в прикладных задачах.

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

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Полное соответствие ответа обучающегося всем перечисленным критериям. Обучающийся демонстрирует высокий уровень владения материалом, ориентируется в предметной области, верно отвечает на все дополнительные вопросы.	Повышенный уровень	Отлично
Ответ на контрольно-измерительный материал не соответствует одному или двум из перечисленных показателей, но обучающийся дает правильные ответы на дополнительные вопросы. Допускаются ошибки при воспроизведении части теоретических положений.	Базовый уровень	Хорошо
Ответ на контрольно-измерительный материал не соответствует любым трём из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. Сформированные знания основных понятий, определений и теорем, изучаемых в курсе, не всегда полное их понимание с затруднениями при воспроизведении.	Пороговый уровень	Удовлетворительно
Ответ на контрольно-измерительный материал не соответствует любым четырём из перечисленных показателей. Обучающийся демонстрирует отрывочные знания (либо их отсутствие) основных понятий, определений и теорем, используемых в курсе.	–	Неудовлетворительно

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 для письменного опроса

Раздел 1. Введение.

1. Оболочка IPython, Jupiter.
2. Горячие клавиши, магические команды.
3. Ошибки и отладка

Раздел 2. Библиотека NumPy

1. Типы данных, массивы, вычисления над массивами.
2. Агрегирование, транслирование, сравнение, булева логика.
3. Индексация, сортировка.
4. Структурированные данные.

Раздел 3. Библиотека Pandas

1. Объекты. Индексация и выборка данных. Операции над данными.

2. Объединение. Агрегирование.
3. Векторизованные операции.
4. Временные ряды.

Раздел 4. Библиотека Matplotlib

1. Графики и диаграммы.
2. Гистограммы.
3. Погрешности.
4. Настройки легенд и шкал цветов.
5. Трехмерные графики.

Раздел 5. Библиотека Scikit-Learn и машинное обучение

1. Гиперпараметры и проверка модели.
2. Проектирование признаков.
3. Наивная байесовская классификация.
4. Линейная регрессия.
5. Метод опорных векторов.
6. Деревья решения и случайные леса.
7. Метод главных компонент.
8. Метод k-средних.
9. Смеси Гауссовых распределений.
10. Ядерная оценка плотности распределения.
11. Конвейер распознавания лиц.

19.3.2 Перечень лабораторных работ

1. Библиотека NumPy
2. Библиотека Pandas
3. Библиотека Matplotlib
4. Библиотека Scikit-Learn
5. Наивная байесовская классификация.
6. Линейная регрессия.
7. Метод опорных векторов.
8. Деревья решения и случайные леса.
9. Метод главных компонент.
10. Метод k-средних.
11. Ядерная оценка плотности распределения.
12. Конвейер распознавания лиц.

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 формах: письменного опроса и контрольных работ. Критерии оценивания приведены выше.

Промежуточная аттестация проводится в соответствии с Положением о промежуточной аттестации обучающихся по программам высшего образования, а также в соответствии с

Положением о балльно-рейтинговой системе контроля знаний на факультете компьютерных наук ВГУ.

Контрольно-измерительные материалы промежуточной аттестации включают в себя теоретические вопросы, позволяющие оценить уровень полученных знаний.

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