

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

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

УТВЕРЖДАЮ

Декан факультета _____

_____/_____/_____

_____ 202_ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Модели и методы машинного обучения»

Направление подготовки 09.04.03 Прикладная информатика

Профиль Инновационные ИТ-проекты и системный инжиниринг

Квалификация выпускника магистр

Нормативный период обучения 2 года и 3 м.

Форма обучения очно-заочная

Год начала подготовки 2024

Автор программы _____ И.О. Фамилия

Заведующий кафедрой

Систем управления и

информационных

технологий в строительстве _____

И.О. Фамилия

Руководитель ОПОП _____

И.О. Фамилия

Воронеж 2024

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

Ознакомление будущих специалистов в области Data Science с процессами, алгоритмы и инструментами, относящимися к основным принципам машинного обучения

1.2. Задачи освоения дисциплины

- Сформировать теоретические знания по основам машинного обучения для построения формальных математических моделей и интерпретации результатов моделирования;
- Выработать умения по практическому применению методов машинного обучения при решении прикладных задач в различных областях;
- Выработать умения и навыки использования библиотек языка Python для разработки систем машинного обучения.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Модели и методы машинного обучения» относится к дисциплинам части, формируемой участниками образовательных отношений блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Модели и методы машинного обучения» направлен на формирование следующих компетенций:

ПК-3 - Владеет инструментами стратегического, компетентностного, календарно-ресурсного планирования, статистического анализа, оценок риска

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-3	Знать современные методы, техники, процессы и инструменты управления требованиями для решения стратегических задач
	Уметь использовать современные методы управления с целью обеспечения конкурентоспособности организаций с помощью анализа макро- и микросреды, учитывая многосценарное регулирование и поведение конкурентов
	Владеть навыками выбора методов и приёмов современного стратегического анализа с учётом его целей, задач и этапов проведения, а также классификации объектов исследования

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Модели и методы машинного обучения» составляет 4 з.е.

Распределение трудоемкости дисциплины по видам занятий

очно-заочная форма обучения

Виды учебной работы	Всего часов	Семестры
		3
Аудиторные занятия (всего)	54	54
В том числе:		
Лекции	18	18
Практические занятия (ПЗ)	36	36
Самостоятельная работа	90	90
Виды промежуточной аттестации - зачет с оценкой	+	+
Общая трудоемкость: академические часы	144	144
зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очно-заочная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Прак зан.	СРС	Всего, час
1	Введение в машинное обучение.	Основные этапы решения задачи анализа данных. Примеры прикладных задач. Основные проблемы машинного обучения: недостаточный объем обучающей выборки, пропуски в данных, переобучение	4	6	14	24
2	Основные определения и постановки задач.	Виды обучения: с учителем, без учителя, с подкреплением. Основные типы задач: задача классификации, задача регрессии, задача кластеризации, задача прогнозирования, задача ранжирования.	4	6	14	24
3	Решение задачи регрессии Решение задачи классификации	Метод наименьших квадратов. Многомерная регрессия, проблема мультиколлинеарности. Регрессия, линейная по параметрам, полиномиальная регрессия. Линейная модель классификации. Логистическая регрессия как линейный классификатор.	4	6	14	24
4	Древовидные модели: деревья решений, случайный лес	Этапы построения дерева решений, выбор критерия точности прогноза. типа ветвления. Метрики ветвления на основе прироста информации. Правила разбиения. Механизм отсечения дерева. Случайный лес. Обучение случайного леса.	2	6	16	24
5	Ансамбли моделей Бэггинг, бустинг, градиентный бустинг	Бэггинг линейных классификаторов. Бустинг. Adaboost для ансамбля из простых деревьев (пней). Модель «мешка слов» для последовательностей из нескольких слов	2	6	16	24
6	Анализ текстовых данных	Продвинутая токенизация, стемминг и лемматизация. Моделирование тем и кластеризация документов. Латентное размещение Дирихле	2	6	16	24
Итого			18	36	90	144

5.2 Перечень лабораторных работ

Не предусмотрено учебным планом

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины не предусматривает выполнение курсового проекта (работы) или контрольной работы.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ПК-3	Знать современные методы, техники, процессы и инструменты управления требованиями для решения стратегических задач	Активное участие в устных опросах на занятиях, правильно отвечает на теоретические вопросы при защите лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Уметь использовать современные методы управления с целью обеспечения конкурентоспособности организаций с помощью анализа макро- и микросреды, учитывая многосценарное регулирование и поведение конкурентов	Выполнение и защита практических работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Владеть навыками выбора методов и приёмов современного стратегического анализа с учётом его целей, задач и этапов проведения, а также классификации объектов исследования	Выполнение и защита практических работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 3 семестре для очной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
ПК-3	Знать современные методы, техники, процессы и инструменты управления требованиями для решения стратегических задач	Тест	Выполнение теста на 90-100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	Уметь использовать современные методы управления с целью обеспечения конкурентоспособности организаций с помощью анализа макро- и микросреды, учитывая многосценарное регулирование и поведение конкурентов	Решение стандартных практических задач	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть навыками выбора методов и приёмов современного стратегического анализа с учётом его целей, задач и этапов проведения, а также классификации объектов исследования	Решение прикладных задач в конкретной предметной области	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типичные контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

Что является ключевым компонентом в работе методов машинного обучения?

- Входные данные.
- Выходные данные.
- Функция потерь.
- Тело модели.
- Оптимизатор.

Какая метрика чаще всего используется для оценки качества бинарной классификации и оценивается в процентах?

- Accuracy.
- R2.
- MSE.
- Precision.

Что такое метод ранней остановки и как он используется для оптимизации модели?

- Метод, который выбирает оптимальное количество эпох для обучения модели.
- Метод, который применяет штрафы за большие веса модели.

в) Метод, который оптимизирует гиперпараметры модели.

г) Метод, который останавливает обучение модели, если ошибка на валидационном наборе начинает увеличиваться.

Что такое гиперпараметры модели?

а) Параметры модели, которые оптимизируются в процессе обучения.

б) Параметры модели, которые используются для оценки её качества.

в) Параметры модели, которые не могут повлиять на процесс обучения модели.

г) Параметры модели, которые определяют её архитектуру и настройки.

Какая компонента модели машинного обучения определяет взаимосвязь между входными и выходными данными?

а) Функция потерь.

б) Тело модели.

в) Оптимизатор.

Что происходит в процессе обучения модели в машинном обучении?

а) Результаты предсказания сравниваются с реальными данными, используя функцию потерь.

б) Всё перечисленное.

Что такое регуляризация модели?

а) Метод, который выбирает оптимальное количество эпох для обучения модели.

б) Метод, который применяет штрафы за большие веса модели.

в) Метод, который оптимизирует гиперпараметры модели.

Какой метод оптимизации традиционно используется для настройки весов в нейронных сетях?

а) Метод оптимизации случайного поиска.

б) Метод оптимизации градиентного спуска.

в) Метод оптимизации Гаусса.

г) Метод оптимизации генетических алгоритмов.

Что делает оптимизатор в машинном обучении?

а) Определяет сложность задачи, которую необходимо решить.

б) Определяет взаимосвязь между входными данными и выходными данными.

в) Используется для оценки качества модели.

г) Минимизирует функцию потерь.

Что такое функция потерь в машинном обучении?

а) Это интерпретируемая функция, которая используется для оценки качества модели и понятная бизнесу.

б) Функция, которая определяет сложность задачи, которую необходимо решить.

в) Функция, которая используется для оценки качества модели во время обучения.

г) Алгоритм, который используется для минимизации функции потерь..

7.2.2 Примерный перечень заданий для решения стандартных задач

Линейная регрессия

- 1) Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
- 2) Обучите, а затем провалидируйте на тестовых данных следующие модели, используя в качестве метрики качества R^2 , предварительно отмасштабировав данные
 - LinearRegression;
 - Lasso с коэффициентом регуляризации, равным 0.01.
- 3) Проанализируйте качество получившихся моделей и сравните количество строго нулевых весов в них.

Логистическая регрессия

- 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 логистической регрессии
- 4) Вычислите значения метрик: recall, precision, F1-мера, AUC-ROC. Постройте ROC-кривую.

7.2.3 Примерный перечень заданий для решения прикладных задач

- *Построение регрессионной модели на основе данных о числе запросов в АТЕ.*
- *Расчёт среднего, медианы, коэффициента асимметрии и эксцесса для данных о динамике цен. Проверка гипотезы о равенстве средних значений темпов роста за заданные годы.*
- *Расчёт распространённости заболевания в заданных АТЕ на основе данных о числе заболеваний. Деление районов на группы с низкой и высокой распространённостью. Сравнение средних значений в полученных группах.*
- *Расчёт темпа прироста населения в заданных АТЕ на основе данных о численности населения. Построение диаграммы рассеяния между полученными показателями. Предположения о форме зависимости между ними.*
- *Прогнозирование уровня первичной заболеваемости БА на основе данных о числе заболеваний в К-ском районе. Получение прогнозного значения на следующий год.*
- *Прогнозирование финансовых потоков в строительные компании.*
- *Поиск схожих рынков по движениям цен и другим факторам,*

определение периодов, когда цены начинают расходиться.

7.2.4 Примерный перечень вопросов для подготовки к зачету

Не предусмотрено учебным планом

7.2.5 Примерный перечень заданий для подготовки к экзамену

1. *Ввод и визуализация одномерных данных в R. Гистограмма, «ящик с усами», их свойства, примеры применения*
2. *Ввод и визуализация двумерных данных в R. Облако точек, его свойства. Облако слов. Примеры применения.*
3. *Интерактивная визуализация данных в R. Библиотека Highcharter. Примеры использования*
4. *Визуализация сетей в виде графов в R. Примеры.*
5. *Построения с использованием географических карт в R. Примеры*
6. *Многомерная визуализация в R. Примеры*
7. *Изменение формата графика в R: масштабирование, агрегирование, иерархия, фильтрация*
8. *Основные классы задач анализа данных и машинного обучения. Примеры.*
9. *Задача классификации: постановка, обзор методов решения, примеры.*
10. *Задача кластеризации: постановка, решение, примеры.*
11. *Линейный регрессионный анализ: постановка задачи, решение, примеры.*
12. *Метод LASSO для выбора переменных в множественной регрессионной модели*
13. *Однофакторная логистическая регрессия: модель, область применения, вычислительные аспекты*
14. *Экспоненциальная регрессия: модель, область применения, вычислительные аспекты*
15. *Множественная логистическая регрессия: модель, область применения, вычислительные аспекты*
16. *Средства оценки качества решения задачи классификации.*
17. *ROC-кривые: основные понятия, построение в R, практическое применение*
18. *Нейронные сети и искусственный интеллект: основные понятия*
19. *Обучение нейронных сетей*
20. *Перцептроны и многослойная архитектура*

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

(Например: Экзамен проводится по тест-билетам, каждый из которых содержит 10 вопросов и задачу. Каждый правильный ответ на вопрос в тесте оценивается 1 баллом, задача оценивается в 10 баллов (5 баллов верное решение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20.

1. *Оценка «Неудовлетворительно» ставится в случае, если студент набрал менее 6 баллов.*

2. Оценка «Удовлетворительно» ставится в случае, если студент набрал от 6 до 10 баллов

3. Оценка «Хорошо» ставится в случае, если студент набрал от 11 до 15 баллов.

4. Оценка «Отлично» ставится, если студент набрал от 16 до 20 баллов.)

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Введение в машинное обучение.	ПК-3	Тест, контрольная работа, защита реферата
2	Основные определения и постановки задач.	ПК-3	Тест, контрольная работа, защита реферата
3	Решение задачи регрессии Решение задачи классификации	ПК-3	Тест, контрольная работа, защита реферата
4	Древовидные модели: деревья решений, случайный лес	ПК-3	Тест, контрольная работа, защита реферата
5	Ансамбли моделей Бэггинг, бустинг, градиентный бустинг	ПК-3	Тест, контрольная работа, защита реферата
6	Анализ текстовых данных	ПК-3	Тест, контрольная работа, защита реферата

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

- Рашка, С. Python и машинное обучение: [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>

- *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>*
- *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>*
- *Плас Дж. Вандер Python для сложных задач: наука о данных и машинное*
- *обучение. — СПб.: Питер, 2018. — 576 с.*
- *Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data.*
- *Python и наука о данных. -СПб.: Питер, 2017. -336 с.*
- *Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017.*
- *-336 с.:*
- *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>*
- *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>*
- *Кук, Д. Машинное обучение с использованием библиотеки H2O [Электронный ресурс] / Д. Кук ; пер. с англ. Огурцова А.Б.. — Электрон. дан. — Москва : ДМК Пресс, 2018. — 250 с. — Режим доступа: <https://e.lanbook.com/book/97353>*

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

- *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-i-2017.compressed-1.pdf

- *Машинное обучение (курс лекций, К.В.Воронцов)*
[http://www.machinelearning.ru/wiki/index.php?title=Машинное обучение](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение)
- *<https://www.kaggle.com/> - онлайн платформа для проектов в области науки о данных*
- *UCI Machine Learning Repository — репозиторий наборов данных для машинного обучения - <http://archive.ics.uci.edu/ml>*
- *Ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. - <http://machinelearning.ru>*
- *Открытый курс машинного обучения <https://habr.com/company/ods/blog/322626/>*

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

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 - Python 3 с подключенными библиотеками (дистрибутив Anaconda)

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Модели и методы машинного обучения» читаются лекции, проводятся практические занятия.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Практические занятия направлены на приобретение практических навыков расчета. Занятия проводятся путем решения конкретных задач в аудитории.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на

	практическом занятии.
Практическое занятие	Конспектирование рекомендуемых источников. Работа с конспектом лекций, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы. Прослушивание аудио- и видеозаписей по заданной теме, выполнение расчетно-графических заданий, решение задач по алгоритму.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоению учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед зачетом с оценкой три дня эффективнее всего использовать для повторения и систематизации материала.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию ОПОП
----------	-----------------------------	----------------------------	--