

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 03.12.2024 22:02:13
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

**Федеральное государственное бюджетное образовательное учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

СЕВЕРО-ЗАПАДНЫЙ ИНСТИТУТ УПРАВЛЕНИЯ – филиал РАНХиГС

ФАКУЛЬТЕТ ЭКОНОМИКИ и ФИНАНСОВ
(наименование структурного подразделения (института/факультета/филиала))
Кафедра экономики
(наименование кафедры)

УТВЕРЖДЕНА

Директор СЗИУ РАНХиГС
Хлутков А.Д.

**ПРОГРАММА МАГИСТРАТУРЫ
Финансовые инструменты в экономике**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ,
реализуемой без применения электронного (онлайн) курса**

Б1.О.05 Машинное обучение
(код и наименование РЦД)

38.04.01 Экономика
(код, наименование направления подготовки)

Очная/ заочная
(формы обучения)

Год набора – 2024

Санкт-Петербург, 2024 г.

Составитель: к.э.н, доцент Котелкин Юрий Владимирович

Заведующий кафедрой экономики, д.э.н., профессор Мисько Олег Николаевич
(наименование кафедры) (ученая степень и(или) ученое звание) (Ф.И.О.)

РПД Б1.О.05 Машинное обучение одобрена на заседании кафедры экономики. Протокол от 19 марта 2024 года № 6.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы.....	4
2. Объем и место дисциплины в структуре образовательной программы.....	5
3. Содержание и структура дисциплины.....	6
4. Материалы текущего контроля успеваемости обучающихся	9
5. Оценочные материалы промежуточной аттестации по дисциплине.....	13
6. Методические материалы для освоения дисциплины	15
7. Учебная литература и ресурсы информационно-телекоммуникационной сети "интернет"	15
7.1. Основная литература.....	15
7.2. Дополнительная литература	16
7.3. Нормативные правовые документы и иная правовая информация.....	16
7.4. Интернет-ресурсы.....	16
7.5. Иные источники.....	17
8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы	17

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1 Дисциплина Б1.О.05 Машинное обучение обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код компонента компетенции	Наименование компонента компетенции
ОПК ОС-3	Способен применять продвинутое инструментальные методы экономического анализа при решении практических и (или) исследовательских задач	ОПК ОС-3.1	Способен применять продвинутое инструментальные методы экономического анализа при решении исследовательских задач
		ОПК ОС-3.2	Способен применять продвинутое инструментальные методы экономического анализа при решении практических задач
ОПК ОС-5	Способен использовать современные информационные технологии и программные средства при решении практических и (или) исследовательских задач	ОПК ОС-5.1	Способен использовать современные информационные технологии и программные средства при решении исследовательских задач
		ОПК ОС-5.2	Способен использовать современные информационные технологии и программные средства при решении практических задач

1.2 В результате освоения дисциплины у студентов должны быть сформированы:

Код компонента компетенции	Результаты обучения
ОПК ОС-3.1 ОПК ОС-3.2	<p><i>на уровне знаний:</i> методы и алгоритмы принятия стратегических решений в экономических системах; общенаучных методов получения эмпирического и теоретического знания;</p> <p><i>на уровне умений:</i> критически оценивать результаты, выявлять перспективные направления исследований;</p> <p><i>на уровне навыков:</i> владения эффективными технологиями в области принятия и исполнения решений.</p>
ОПК ОС-5.1 ОПК ОС-5.2	<p>сформировать теоретические знания по основам машинного обучения для построения формальных математических моделей и интерпретации результатов моделирования;</p> <p>выработать умения по практическому применению методов машинного обучения при решении прикладных задач в различных областях;</p> <p>выработать навыки использования библиотек языка Python для разработки систем машинного обучения</p>

2. Объем и место дисциплины в структуре образовательной программы

Учебная дисциплина «Машинное обучение» включена в состав обязательных дисциплин образовательной программы.

Общая трудоемкость дисциплины составляет 4 зачетных единиц 144 академ. часа.

Вид работы	Трудоемкость очно/заочно (в акад. часах)
Общая трудоемкость	144
Контактная работа	66/18
Лекции	32/8
Практические занятия	32/8
Консультация	2/2
Самостоятельная работа	42/117
Контроль самостоятельной работы	36/9
Форма промежуточной аттестации	Экзамен

Дисциплина «Машинное обучение» является теоретическим и практико-ориентированным курсом, в процессе изучения которого магистранты знакомятся с основными методами и инструментами оценки и прогнозирования финансово-экономических рисков.

Доступ к системе дистанционных образовательных технологий осуществляется каждым обучающимся самостоятельно с любого устройства на портале: <https://lms.ranepa.ru/>. Пароль и логин к личному кабинету / профилю предоставляется студенту в деканате.

Все формы текущего контроля, проводимые в системе дистанционного обучения, оцениваются в системе дистанционного обучения. Доступ к видео и материалам лекций предоставляется в течение всего семестра. Доступ к каждому виду работ и количество попыток на выполнение задания предоставляется на ограниченное время согласно регламенту дисциплины, опубликованному в СДО. Преподаватель оценивает выполненные обучающимся работы не позднее 10 рабочих дней после окончания срока выполнения.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Очная форма обучения

№ п/п	Наименование тем	Объем дисциплины, час.					Форма текущего контроля успеваемости**, промежуточной аттестации***	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					СР
			Л/ДОТ	ЛР/ДОТ	ПЗ/ДОТ	КСР		
Тема 1	Введение в машинное обучение. Основные определения и постановки задач.	13	4		4		5	РЗ* Т**
Тема 2	Обзор основных необходимых библиотек языка Python	14	4		4		6	РЗ* Т**
Тема 3	Построение и отбор признаков	13	4		4		5	РЗ* Т**
Тема 4	Решение задачи регрессии	13	4		4		5	РЗ* Т**
Тема 5	Решение задачи классификации.	13	4		4		5	РЗ* Т**
Тема 6	Древовидные модели: деревья решений, случайный лес	13	4		4		5	РЗ* Т**
Тема 7	Ансамбли моделей Бэггинг, бустинг, градиентный бустинг	14	4		4		6	
Тема 8	Анализ текстовых данных	13	4		4		5	РЗ* Т**
Консультации на промежуточную аттестацию		2						
Промежуточная аттестация		36						экзамен
Всего:		144	32		32	36	42	

РЗ* – решение задач, Т** - тестирование

заочная форма обучения

№ п/п	Наименование тем	Объем дисциплины, час.					Форма текущего контроля успеваемости**, промежуточной аттестации***	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					СР
			Л/ДОТ	ЛР/ДОТ	ПЗ/ДОТ	КСР		
Тема 1	Введение в машинное обучение. Основные определения и постановки задач.	16	1		1		14	РЗ* Т**
Тема 2	Обзор основных необходимых библиотек языка Python	17	1		1		15	РЗ* Т**
Тема 3	Построение и отбор признаков	17	1		1		15	РЗ* Т**
Тема 4	Решение задачи регрессии	17	1		1		15	РЗ* Т**

Тема 5	Решение задачи классификации.	16	1		1		14	РЗ* Т**
Тема 6	Древовидные модели: деревья решений, случайный лес	17	1		1		15	РЗ* Т**
Тема 7	Ансамбли моделей Бэггинг, бустинг, градиентный бустинг	17	1		1		15	
Тема 8	Анализ текстовых данных	16	1		1		14	РЗ* Т**
Консультации на промежуточную аттестацию		2						
Промежуточная аттестация		9						экзамен
Всего:		144	8		8	9	117	

РЗ* – решение задач, Т** - тестирование

3.2. Содержание дисциплины

Тема 1. Введение в машинное обучение. Основные определения и постановки задач.

Лекция: Основные этапы решения задачи анализа данных. Примеры прикладных задач. Виды обучения: с учителем, без учителя, с подкреплением. Основные типы задач: задача классификации, задача регрессии, задача кластеризации, задача прогнозирования, задача ранжирования. Основные проблемы машинного обучения: недостаточный объем обучающей выборки, пропуски в данных, переобучение.

Тема 2. Обзор основных необходимых библиотек языка Python

Практические занятия: Библиотека NumPy для оптимизированных вычислений над массивами данных. Введение в массивы библиотеки NumPy. Выполнение вычислений над массивами библиотеки NumPy, универсальные функции. Операции над данными в библиотеке Pandas. Обработка отсутствующих данных. Агрегирование и группировка. Визуализация с помощью библиотеки Matplotlib. Линейные графики, диаграммы рассеяния, гистограммы, трехмерные графики. Знакомство с библиотекой машинного обучения Scikit-Learn. Гиперпараметры и проверка качества модели

Тема 3. Построение и отбор признаков

Практические занятия: Извлечение признаков (Feature Extraction). Преобразования признаков (Feature transformations): кодирование нечисловых данных, нормировка и калибровка, заполнение пропусков. Выбор признаков (Feature selection): статистические подходы, визуализация, отбор с использованием моделей

Тема 4. Решение задачи регрессии

Лекция: Метод наименьших квадратов. Измерение ошибки в задачах регрессии (MSE , $RMSE$, MAE , R^2). Многомерная регрессия, проблема мультиколлинеарности. Регрессия, линейная по параметрам, полиномиальная регрессия. Решение проблемы переобучения: L1- регуляризация (Lasso), L2- Регуляризация (гребневая регрессия), эластичная сеть. Настройка гиперпараметров алгоритма с помощью n-кратной перекрестной проверки

Практические занятия: Разбор примера построения модели линейной регрессии для задачи предсказания велосипедного трафика. Отбор и кодирование признаков. Визуальное сравнение общего и предсказанного моделью трафика. Проверка качества. Построение модели линейной регрессии с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели.

Тема 5. Решение задачи классификации.

Лекция: Линейная модель классификации. Логистическая регрессия как линейный классификатор. Функция потерь (ошибок классификации). Логистическая функция потерь с учетом L2- регуляризации. Использование полиномиальных признаков для нелинейного разделения. Confusion matrix (матрица ошибок классификации). Метрики качества классификации: accuracy (доля правильных ответов), precision (точность), recall (полнота), F1- мера. AUC-ROC – площадь под кривой ошибок. Метрическая классификация - метод ближайших соседей (kNN). Использование наивной байесовской модели для классификации

Практические занятия: Разбор примера построения модели логистической регрессии для задачи предсказания оттока клиентов мобильного оператора. Отбор и кодирование признаков. Проверка качества модели с помощью перекрёстной проверки. Построение модели логистической регрессии с помощью библиотеки Scikit-Learn. Анализ качества построенной модели

Тема 6. Древовидные модели: деревья решений, случайный лес

Лекция: Этапы построения дерева решений, выбор критерия точности прогноза. типа ветвления. Метрики ветвления на основе прироста информации (алгоритм ID3), нормализованного прироста информации (алгоритм C4.5), индекса Джини (алгоритм CART). Правила разбиения. Механизм отсечения дерева. Критерии останова алгоритма (минимальное число объектов, при котором выполняется расщепление, минимальное число объектов в листьях, максимальная глубина деревьев. Переобучение решающих деревьев. Случайный лес. Обучение случайного леса. Достоинства и недостатки случайного леса

Практические занятия: Разбор примера построения модели дерева решений для задачи предсказания исхода футбольного матча. Анализ деревьев, полученных при использовании различных метрик. Построение модели случайного леса на примере задачи кредитного скоринга. Кодирование признаков и заполнение пропущенных данных. Построение моделей деревьев решений и случайного леса с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели.

Тема 7. Ансамбли моделей Бэггинг, бустинг, градиентный бустинг

Лекция: Бэггинг, случайный лес как пример бэггинга. Бэггинг линейных классификаторов. Бустинг. Adaboost для ансамбля из простых деревьев (пней). Сравнение результатов бустинга для слабых и сильных моделей. Градиентный бустинг. Градиентный бустинг в задаче регрессии. Градиентный бустинг в задаче классификации. Градиентный бустинг над деревьями.

Практические занятия: Разбор примера построения модели градиентного бустинга для задачи распознавания рукописных цифр из библиотеки MNIST. Построение модели

градиентного бустинга с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели.

Тема 8. Анализ текстовых данных

Лекция: Представление текстовых данных в виде «мешка слов». Стопслова. Масштабирование данных с помощью tf-idf. Модель «мешка слов» для последовательностей из нескольких слов (nграмм) Продвинутая токенизация, стемминг и лемматизация Моделирование тем и кластеризация документов. Латентное размещение Дирихле

Практические занятия: Разбор примера построения модели анализа текстовых данных для задачи определения тональности киноотзывов. Построение модели анализа текстов с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели.

4. Материалы текущего контроля успеваемости обучающихся

4.1. В ходе реализации дисциплины Б1.0.05 «Машинное обучение» используются следующие методы текущего контроля успеваемости обучающихся:

- при проведении занятий лекционного типа: опрос на практическом занятии.
- при проведении практических занятий: решение задач, тестирование, опрос на практическом занятии.
- при контроле результатов самостоятельной работы студентов: контрольная работа, опрос на практическом занятии.

Тема 1	Введение в машинное обучение. Основные определения и постановки задач.	РЗ* Т**
Тема 2	Обзор основных необходимых библиотек языка Python	РЗ* Т**
Тема 3	Построение и отбор признаков	РЗ* Т**
Тема 4	Решение задачи регрессии	РЗ* Т**
Тема 5	Решение задачи классификации.	РЗ* Т**
Тема 6	Древовидные модели: деревья решений, случайный лес	РЗ* Т**
Тема 7	Ансамбли моделей Бэггинг, бустинг, градиентный бустинг	
Тема 8	Анализ текстовых данных	РЗ* Т**

4. 2. Материалы текущего контроля успеваемости обучающихся.

4.2. Типовые материалы текущего контроля успеваемости обучающихся

4.2.1. Типовые материалы текущего контроля успеваемости обучающихся к теме 1

Задание 1: Что такое линейная регрессия?

Вариант 1 это функциональная зависимость, которая позволяет прогнозировать изменения непрерывных числовых параметров;

Вариант 2 модель зависимости непрерывной переменной y от объясняющих ее факторов, в которой функция зависимости является линейной

Вариант 3 модель зависимости дискретной переменной y от объясняющих ее факторов, в которой функция зависимости является линейной

4.2.2. Типовые материалы текущего контроля успеваемости обучающихся к теме 2

Задание 2: Основная характеристика задачи бинарной классификации:

Вариант 1 классификация осуществляется по одному признаку

Вариант 2 зависимая переменная может принимать только два значения

Вариант 3 классификация осуществляется по двум признакам

Задание 3: Классификация относится к стратегии:

Вариант 1 обучения без учителя

Вариант 2 обучения с учителем

Вариант 3 оба ответа неверны

4.2.3. Типовые материалы текущего контроля успеваемости обучающихся к теме 3

Задание 4: Явление переобучения характеризуется

Вариант 1 чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению

Вариант 2 возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры модели

Вариант 3 возникновением, в случае слишком долгого обучения, слишком сложной структуры модели

Задание 5: Задача классификации сводится к ...

Вариант 1 нахождению частых зависимостей между объектами или событиями;

Вариант 2 определению класса объекта по его характеристикам;

Вариант 3 определению по известным характеристикам объекта значение некоторого его параметра;

Вариант 4 поиску независимых групп и их характеристик в всем множестве анализируемых данных.

4.2.4. Типовые материалы текущего контроля успеваемости обучающихся к теме 4

Ответьте на вопросы о данных по авиарейсам в США за 4 месяца. По ссылке расположены Данные и их описание

1) Считайте выборку из файла при помощи функции `pd.read_csv` и ответьте на следующие вопросы:

- Имеются ли в данных пропущенные значения?
- Сколько всего пропущенных элементов в таблице "объект-признак"?
- Сколько объектов имеют хотя бы один пропуск?
- Сколько признаков имеют хотя бы одно пропущенное значение?

2) Преобразуйте каждый признак `FeatureName` из указанных в пару новых признаков `FeatureName_Hour`, `FeatureName_Minute`, разделив каждое из значений на часы и минуты. Не забудьте при этом исключить исходный признак из выборки. В случае, если значение

признака отсутствует, значения двух новых признаков, его заменяющих, также должны отсутствовать.

3) Некоторые из признаков, отличных от целевой переменной, могут оказывать чересчур значимое влияние на прогноз, поскольку по своему смыслу содержат большую долю информации о значении целевой переменной. Изучите описание датасета и исключите признаки, сильно коррелирующие с ответами. Ваш выбор признаков для исключения из выборки обоснуйте.

4) Приведите данные к виду, пригодному для обучения линейных моделей. Для этого вещественные признаки надо отмасштабировать, а категориальные — привести к числовому виду. Также надо устранить пропуски в данных. Реализуйте функцию `transform_data`, которая принимает на вход `DataFrame` с признаками и выполняет следующие шаги:

- Замена пропущенных значений на нули для вещественных признаков и на строки 'nan' для категориальных.

- Масштабирование вещественных признаков с помощью `StandardScaler`.

- One-hot-кодирование категориальных признаков с помощью `DictVectorizer` или функции `pd.get_dummies`. Метод должен возвращать преобразованный `DataFrame`, который должна состоять из масштабированных вещественных признаков и закодированных категориальных (исходные признаки должны быть исключены из выборки).

5) Разбейте выборку и вектор целевой переменной на обучение и контроль в отношении 70/30 (для этого можно использовать функцию `train_test_split`).

6) Обучите линейную регрессию на 1000 объектах из обучающей выборки и выведите значения MSE и R^2 и на этой подвыборке и контрольной выборке. Проинтерпретируйте полученный результат — насколько качественные прогнозы строит полученная модель? Какие проблемы наблюдаются в модели?

7) Обучите линейные регрессии с L1- и L2-регуляризатором, подобрав лучшее значение параметра регуляризации при помощи кросс-валидации. Выведите значения и на обучающей и контрольной выборках. Удалось ли решить указанные вами ранее проблемы?

4.2.5. Типовые материалы текущего контроля успеваемости обучающихся к теме 5

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.

4.2.6. Типовые материалы текущего контроля успеваемости обучающихся к теме 6

Задание 6: Множество примеров, используемое для конструирования модели, называется...

Вариант 1 обучающим множеством

Вариант 2 тестовым множеством

Вариант 3 валидационным множеством

4.2.7. Типовые материалы текущего контроля успеваемости обучающихся к теме 7

Задание 7: В задаче определения спама recall - это...

Вариант 1 доля спама среди писем, распознанных алгоритмом как спам

Вариант 2 доля не спама среди писем, распознанных алгоритмом как спам

Вариант 3 доля писем, распознанных алгоритмом как спам, среди спама

Вариант 4 доля писем, распознанных алгоритмом как не спам, среди спама

4.2.8. Типовые материалы текущего контроля успеваемости обучающихся к теме 8

Практическое задание посвящено работе с текстовыми данными и категориальными признаками и задачам бинарной классификации. В рамках данного задания нужно решить задачу бинарной классификации для предсказания уровня заработной платы по тексту объявления о вакансии на примере набора данных с соревнования на Kaggle. Данные доступны по ссылке.

1) Разбейте получившуюся выборку на обучающую и контрольную в соотношении 70/30

2) Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, осуществив векторизацию получившихся текстов при помощи CountVectorizer, обучив его на обучающей выборке и применив на тестовой.

3) Обучите логистическую регрессию из модуля sklearn с параметрами по умолчанию на обучающей выборке:

4) Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на тестовой выборке.

5) Отсортируйте веса признаков для модели. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат.

6) Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, вычислив вектор tf-idf для каждого объекта помощи TfidfVectorizer, обучив его на обучающей выборке и применив на тестовой.

7) Заново обучите модель

8) Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на контрольной выборке..

9). Сравните значения метрик из п. 8 со значениями, полученными в п. 4, и сравните соответствующие модели по качеству из работы.

10). Отсортируйте веса признаков для модели логистической регрессии из scikit-learn, полученной в п. 7. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат.

5. Оценочные материалы промежуточной аттестации по дисциплине

5.1 Экзамен проводится с применением следующих методов (средств): в форме контрольной работы по билетам или в виде теста. На экзамен выносится материал в объеме, предусмотренном рабочей программой учебной дисциплины.

При реализации промежуточной аттестации в ЭО/ДОТ могут быть использованы следующие формы:

1. Устно в ДОТ - в форме устного ответа на теоретические вопросы и решения задачи (кейса).

2. Письменно в СДО с прокторингом - в форме письменного ответа на теоретические вопросы и решения задачи (кейса).

3. Тестирование в СДО с прокторингом

5.2. Оценочные материалы промежуточной аттестации

Компонент компетенции	Промежуточный/ключевой индикатор оценивания	Критерий оценивания
ОПК ОС-3.1 Способен применять продвинутое инструментальные методы экономического анализа при решении исследовательских задач	Применяет продвинутое инструментальные методы экономического анализа при решении исследовательских задач	Умение применять продвинутое инструментальные методы экономического анализа при решении исследовательских задач
ОПК ОС-3.2 Способен применять продвинутое инструментальные методы экономического анализа при решении практических задач	Применяет продвинутое инструментальные методы экономического анализа при решении практических задач	Умение применять продвинутое инструментальные методы экономического анализа при решении практических задач
ОПК ОС-5.1 Способен использовать современные информационные технологии и программные средства при решении исследовательских задач	Использует современные информационные технологии и программные средства при решении исследовательских задач	Умение использовать современные информационные технологии и программные средства при решении исследовательских задач
ОПК ОС-5.2 Способен использовать современные информационные технологии и программные средства при решении практических задач	Использует современные информационные технологии и программные средства при решении практических задач	Умение использовать современные информационные технологии и программные средства при решении практических задач

Типовые оценочные материалы промежуточной аттестации

Перечень вопросов к экзамену

1. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
2. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
3. Метрики качества алгоритмов регрессии и классификации.
4. Линейная регрессия. Простая многомерная регрессия. Регрессия с полиномиальными признаками. Методы регуляризации: Ridge, Lasso, ElasticNet.
5. Логистическая регрессия.
6. Деревья решений. Методы построения деревьев. Их регуляризация.
7. Композиции алгоритмов. Разложение ошибки на смещение и разброс.
8. Случайный лес, его особенности.
9. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.
10. Анализ текстов. Масштабирование данных с помощью tf-idf. Модель «мешка слов» для n-грамм.

Шкала оценивания

Для оценивания результатов обучения на экзамене используются следующие показатели:

- 1) знание учебного материала и владение понятийным аппаратом теории машинного обучения;
- 2) умение анализировать многомерные данные и преодолевать вычислительные проблемы, связанные с высокой размерностью данных;
- 3) умение применять методы машинного обучения при решении задач в различных прикладных областях;
- 4) владение навыками использования библиотек языка Python для построения систем, обучающихся по прецедентам
- 5) владение навыками построения и проверки качества моделей машинного обучения;
- 6) владение навыками интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов

На оценку «Отлично» обучающийся в полной мере владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), сдал все практические и лабораторные работы, среднее количество правильных ответов на вопросы тестов превышает 80%.

На оценку «Хорошо» обучающийся владеет понятийным аппаратом данной области науки (теоретическими основами дисциплины), но не сдал одну практическую или лабораторную работу, среднее количество правильных ответов на вопросы тестов находится в диапазоне 70-80%.

На оценку «Удовлетворительно» обучающийся демонстрирует неуверенное владение понятийным аппаратом данной области науки (теоретическими основами дисциплины), не сдал две практических или лабораторных работы, среднее количество правильных ответов на вопросы тестов находится в диапазоне 60-70%.

На оценку «Неудовлетворительно» обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки, не сдал более двух практических или лабораторных работы, среднее количество правильных ответов на вопросы тестов менее 70%.

6. Методические материалы для освоения дисциплины

Основными видами занятий магистрантов являются: лекционный курс и семинарские (практические) занятия на которых магистры вместе с преподавателем обсуждают выполненные задания.

Теоретические занятия (интерактивные лекции) организуются по потокам. Практические занятия организуются по группам с использованием активных и интерактивных форм проведения занятий (в диалоговом режиме, дискуссий, деловых и ролевых игр, разбор конкретных ситуаций «кейс-стади», групповых дискуссий) в сочетании с внеаудиторной работой с целью формирования и развития компетенций обучающихся. Также в качестве варианта проведения практического занятия используется форма обсуждения подготовленных магистрантами вопросов по заданным темам.

Изучение данной дисциплины предусматривает также самостоятельную работу магистранта. Выполнение самостоятельной работы предполагает: качественную подготовку ко всем видам учебных занятий; реферирование и аннотирование указанных преподавателем источников и литературы; систематический просмотр периодических изданий целью выявления публикаций в области изучаемой проблематики; изучение учебной литературы; использование интернет-ресурсов; подготовку докладов-презентаций по отдельным темам дисциплины.

В процессе самостоятельной подготовки при освоении дисциплины необходимо изучить основную литературу, затем – дополнительную. Именно знакомство с дополнительной литературой, значительная часть которой существует как в печатном, так и электронном виде, способствует более глубокому освоению изученного материала. Литературу можно найти в указанных выше источниках, сети Интернет.

Выступления на практических занятиях могут быть представлены в виде реферата, доклада или сообщения. Любое из них должно содержать план или постановку задачи, изложение материала и выводы. В каждом выступлении необходимо выделять главную мысль («стержневой вопрос»). Выступления должны носить научный, логичный, аргументированный, конкретный и профессиональный характер, быть убедительными.

7. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет"

7.1. Основная литература

1. 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>

2. 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>
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>

7.2. Дополнительная литература

4. Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.
5. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. -СПб.: Питер, 2017. -336 с.
6. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017. -336 с.:
7. 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>
- 8 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>
- 9 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> бучение. — СПб.: Питер, 2018. — 576 с.
10. 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-mashinnoebuchenie-s-pomoshhyu-python.-rukovodstvo-dlya-specialistov-po-rabote-s-dannymi2017.compressed-1.pdf>

7.3. Нормативные правовые документы и иная правовая информация

Не используются

7.4. Интернет-ресурсы.

1. UCI Machine Learning Repository — репозиторий наборов данных для машинного обучения - <http://archive.ics.uci.edu/ml/>
2. Ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. - <http://machinelearning.ru>
3. Открытый курс машинного обучения <https://habr.com/company/ods/blog/322626/>

Электронно-образовательные ресурсы на сайте научной библиотеки СЗИУ РАНХиГС

1. Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»
http://www.nwapa.spb.ru/index.php?page_id=76
2. Статьи из журналов и статистических изданий Ист Вью
http://www.nwapa.spb.ru/index.php?page_id=76
3. Энциклопедии и справочники компании Рубрикон
4. Polpred.com Обзор СМИ.
5. EBSCO Publishing - доступ к журналам таких издательств как Blackwell publishers, Springer, Elsevier, Harvard business school, Taylor and Francis, Academy of Management, Transaction publishers, American institute of physics, University of california press и многие другие.
6. Мировое издательство Emerald eJournals Premier - электронное собрание рецензируемых журналов по всем основным дисциплинам менеджмента
7. Архив научных журналов 2011 Cambridge Journals Digital Archive Complete Collection издательства Cambridge University Press:
<http://journals.cambridge.org/action/displaySpecialPage?pageId=3092&archive=3092>
8. Международное издательство SAGE Publications (штаб-квартиры в США, Великобритании (Лондон), Индии)
9. Американское издательство Annual Reviews
10. Oxford Journals Archive - архив политематических научных журналов издательства Oxford University Press.
11. T&F 2011 Journal Archives Collection - архив научных журналов издательства Taylor and Francis.
12. The American Association for the Advancement of Science (AAAS) - цифровой архив статей журнала Science.
13. Nature journal Digital archive - цифровой архив журнала Nature издательства Nature Publishing Group.

7.5. Иные источники

Не используются

8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Характеристики аудиторий (помещений, мест) для проведения занятий

Для проведения занятий необходимы стандартно оборудованные учебные кабинеты и компьютерные классы соответствующие санитарным и строительным нормам и правилам.

№ п/п	Наименование
1.	Специализированные залы для проведения лекций:

2.	Специализированная мебель и оргсредства: аудитории
3.	Технические средства обучения: Персональные компьютеры; компьютерные проекторы; звуковые динамики; программные средства, обеспечивающие просмотр видеофайлов в форматах AVI, MPEG-4, DivX, RMVB, WMV.

На практических занятиях используется следующее программное обеспечение:

- программы, обеспечивающие доступ в сеть Интернет (например, «Google chrome»);
- программы, демонстрации видео материалов (например, проигрыватель « Windows Media Player»);
- программы для демонстрации и создания презентаций (например, «Microsoft Power Point»);
- пакеты прикладных программ SPSS/PC+, СТАТИСТИКА,
- программные комплексы Word, ТЕСТУНИВЕРСАЛ,
- системы дистанционного обучения «Прометей» и WebSoft,
- компьютерная система «e-Learning-IDOX», СУБД MS Access,
- правовые базы данных «Консультант+», «Гарант», «Кодекс», «Эталон»