

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ»

Саровский физико-технический институт -

филиал федерального государственного автономного образовательного учреждения высшего
образования «Национальный исследовательский ядерный университет «МИФИ»
(СарФТИ НИЯУ МИФИ)

ФИЗИКО-ТЕХНИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра «Цифровых технологий»

УТВЕРЖДАЮ

Декан ФТФ, член корр. РАН, д.ф-м.н.

_____ **А.К. Чернышев**

« ____ » _____ **2023 г.**

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Машинное обучение и анализ данных

наименование дисциплины

Направление подготовки (специальность)	09.03.02 Информационные системы и технологии
Наименование образовательной программы	Инновационные технологии комплексной автоматизации и сквозного управления жизненным циклом
Квалификация (степень) выпускника	бакалавр
Форма обучения	очная

Программа одобрена на заседании кафедры	Зав. кафедрой ЦТ
	_____ О.В. Кривошеев
протокол № от 2023 г.	« ____ » _____ 2023 г.

г. Саров, 2023 г.

Программа переутверждена на 202____/202____ учебный год с изменениями в соответствии с семестровыми учебными планами академических групп ФИТЭ, ФТФ на 202____/202____ учебный год.

Заведующий кафедрой ЦТ

О.В. Кривошеев

Программа переутверждена на 202____/202____ учебный год с изменениями в соответствии с семестровыми учебными планами академических групп ФИТЭ, ФТФ на 202____/202____ учебный год.

Заведующий кафедрой ЦТ

О.В. Кривошеев

Программа переутверждена на 202____/202____ учебный год с изменениями в соответствии с семестровыми учебными планами академических групп ФИТЭ, ФТФ на 202____/202____ учебный год.

Заведующий кафедрой ЦТ

О.В. Кривошеев

Программа переутверждена на 202____/202____ учебный год с изменениями в соответствии с семестровыми учебными планами академических групп ФИТЭ, ФТФ на 202____/202____ учебный год.

Заведующий кафедрой ЦТ

О.В. Кривошеев

Семестр	В форме практической подготовки	Трудоемкость, кред.	Общий объем курса, час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	СРС, час.	КР/ КП	Форма(ы) контроля, экз./зач./ЗсО/	Интерактивные часы
5	16	4	72	32	-	32	44	-	Экз	16
ИТОГО	16	4	72	32	-	32	44	-	36	16

АННОТАЦИЯ

В рамках данного курса предусмотрено изучение дисциплины теоретических основ и применение на практике полученных знаний и навыков исследования данных и разработки математических моделей, методов и алгоритмов анализа данных.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Цель освоения дисциплины «Машинное обучение и анализ данных» формирует у будущих специалистов модели, алгоритмы, методы машинного обучения, инструментарию, а также позволяет приобрести навыки исследования данных и разработки математических моделей, методов и алгоритмов анализа данных.

Задачи преподавания дисциплины — это изучение:

- Методов анализа данных;
- Методов сбора данных;
- Методов обработки данных;
- Алгоритмы машинного обучения;
- Методов визуализации данных;
- Освоение современных аналитических инструментов.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

«Машинное обучение и анализ данных» относится к части, формируемой участниками образовательных отношений ООП «Инновационные технологии комплексной автоматизации и сквозного управления жизненным циклом» ОС НИЯУ МИФИ по направлению 09.03.02 «Информационные системы и технологии».

Учебная дисциплина базируется на знаниях, полученных в ходе изучения следующих дисциплин: математический анализ, алгебра и аналитическая геометрия, теория вероятности, математическая статистика, информатика, программирование, численные методы, программирование на языке Python.

Изучением дисциплины обеспечивает формирование у специалистов представление об алгоритмах, математических моделях и программной реализации.

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Ожидается, что в результате освоения дисциплины студент приобретет следующие компетенции:

Профессиональные компетенции (ПК)

в соответствии с задачами и объектами (областями знаний) профессиональной деятельности:

Задача профессиональной деятельности (ЗПД)	Объект или область знания	Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции
Тип задачи профессиональной деятельности: производственно-технологический			
разработка и внедрение технологий разработки объектов профессиональной деятельности в различных областях и сферах деятельности	информационные процессы, технологии, системы и сети, их инструментальное (программное, техническое, организационное) обеспечение, способы и методы проектирования, отладки, производства и эксплуатации информационных технологий и систем в различных областях и сферах деятельности.	ПК-5 Способен осуществлять моделирование процессов и систем на основе системного анализа предметной области Профессиональный стандарт 06.015 Специалист по информационным системам	З-ПК-5 Знать: основные принципы системного подхода; методы моделирования процессов и систем У-ПК-5 Уметь: проводить анализ предметной области и осуществлять ее формальное представление в виде модели В-ПК-5 Владеть: инструментальными средствами моделирования
Тип задачи профессиональной деятельности: проектный			
проектирование базовых и прикладных информационных технологий	информационные процессы, технологии, системы и сети, их инструментальное (программное, техническое, организационное) обеспечение, способы и методы проектирования, отладки, производства и эксплуатации информационных технологий и систем в различных областях и сферах деятельности.	ПК-14 Способен к проектированию базовых и прикладных информационных технологий Профессиональный стандарт 06.001 Программист	З-ПК-14 Знать: методы управления данными; технологии программирования У-ПК-14 Уметь: разрабатывать приложения и сервисы; создавать программный код; разрабатывать базы данных и организовывать доступ к ним В-ПК-14 Владеть: языками программирования

	сферах деятельности.		ния; форматами представления данных и СУБД
Тип задачи профессиональной деятельности: производственно-технологический			
разработка и внедрение технологий разработки объектов профессиональной деятельности в различных областях и сферах деятельности	информационные процессы, технологии, системы и сети, их инструментальное (программное, техническое, организационное) обеспечение, способы и методы проектирования, отладки, производства и эксплуатации информационных технологий и систем в различных областях и сферах деятельности.	ПК-2.2 Способен проводить подготовку документации по менеджменту качества информационных систем Профессиональный стандарт 06.015 Специалист по информационным системам	З-ПК-2.2 знать современный отечественный и зарубежный опыт в профессиональной деятельности У-ПК-2.2 уметь анализировать исходные данные, планировать работы В-ПК-2.2 владеть навыками подготовки документации по менеджменту качества информационных систем

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ*

№ п/п	Наименование раздела /темы дисциплины	№ недели	Виды учебной работы					
			Лекции	Практ. заня- тия/ семинары	Лаб. работы	СРС	Текущий кон- троль (форма)*	Максимальный балл (см. п. 5.3)
			32	-	32	44		
Семестр 5								
1.1.	Введение в машинное обучение (МО) и анализ данных (АД)	1	2			4	-	-
1.2	Основы языка программирования Python	2	2		4	4	УО	2
1.3	Сведения из математической статистики, необходимые для МО и АД	3	2			4	УО	1
1.4	А/В гипотезы	4	2		4	4	УО	1
1.5	Алгоритмы ассоциации	5	4		4	4	УО	1
Рубежный контроль		7	Тест					10
2.1	Алгоритмы кластеризации	6	2		4	4	УО	2
2.2	Линейная регрессия	7	2		-	4	УО	2
2.3	Алгоритмы классификации (статистические методы)	8	2		4	4	УО	2
2.4	Алгоритмы классификации (методы МО)	9	2		4	4	УО	2
2.5	Повышение точности классификаторов	10	4		4	4	УО	2

№ п/п	Наименование раздела /темы дисциплины	№ недели	Виды учебной работы					
			Лекции	Практ. занятия/ семинары	Лаб. работы	СРС	Текущий контроль (форма)*	Максимальный балл (см. п. 5.3)
			32	-	32	44		
2.6	Оценка эффективности и точности моделей	11	2		4	4	УО	2
Рубежный контроль		16	Тест					10
Промежуточная аттестация			зачет				36	0 - 50
Посещаемость								5
Итого:			32	-	32	44		100

*Сокращение наименований форм текущего, рубежного и промежуточного контроля:

УО – устный опрос

Тест – тестирование (письменный опрос)

Э/Зач/ЗсО – экзамен/зачет/зачет с оценкой и др.

4.2. Содержание дисциплины, структурированное по разделам (темам)

Лекционный курс

№	Наименование раздела /темы дисциплины	Содержание
1.	Введение в машинное обучение (МО) и анализ данных (АД)	Бизнес-аналитика, решаемые задачи, большие данные. Технологии обработки больших данных. Введение в алгоритмы МО и АД.
2.	Основы языка программирования Python	Обзор языков и систем программирования, используемых в машинном обучении. Обязательные библиотеки для машинного обучения NumPy, SciPy, Matplotlib, Pandas. Типы данных, включая массивы библиотеки NumPy, проверка типа данных. Операторы if, for. Построение графиков и гистограмм.
3.	Сведения из математической статистики, необходимые для МО и АД	Случайные величины, функция распределения и функция плотности, примеры законов распределения из задач МО и АД - нормальный, биномиальный, Пуассона. Параметры распределений: константы и/или случайные величины. Доверительные интервалы. Бутстреп и его возможности. Понятие статистической гипотезы, ошибки I и II рода.
4.	А/В гипотезы	Примеры задач бизнес-аналитики, сводящиеся к А/В гипотезам, понятия и примеры метрик. Математическая (статистическая) постановка задачи, ошибки I и II рода. Алгоритм аналитического решения для нормальных распределений. Алгоритм численного (бутстреп) решения для произвольного (непараметрического) распределения и любой метрики. Сравнение области применения каждого алгоритма.
5.	Алгоритмы ассоциации	Примеры задач бизнес-аналитики, сводящиеся к нахождению ассоциаций. Алгоритм Apriori. Алгоритм иерархических ассоциативных правил. Алгоритм последовательных шаблонов. Сравнение и области применения каждого алгоритма.
6.	Алгоритмы кластеризации	Примеры задач бизнес-аналитики, сводящиеся к кластеризации. Алгоритм к-средних (k-means), метрики –линейная

		и квадратичная, области применения каждой метрики. Алгоритм G-mean или сеть Кохонена (один из вариантов нейронной сети). Сравнение и области применения каждого алгоритма.
7.	Линейная регрессия	Примеры задач бизнес-аналитики, сводящиеся к линейной регрессии. Идея линейной регрессии, требования к исходным данным, оценка соответствия линейной модели данным (коэффициент корреляции, t- и F-критерии). Построение линейной регрессии средствами Python (scipy). Многомерная линейная регрессия, понижение размерности. Ограничения и область применения линейной регрессии в задачах бизнес аналитики.
8.	Алгоритмы классификации (статистические методы, методы машинного обучения)	Примеры задач бизнес-аналитики, сводящиеся к классификации методами опорных векторов и логистической регрессией. Алгоритм метода опорных векторов (SVM). Логистическая регрессия. Сравнение и области применения каждого алгоритма. Примеры задач бизнес-аналитики, сводящиеся к классификации алгоритмом наивного байесовского классификатора. Алгоритм байесовского классификатора. Вводные понятия решения задачи пакетом PyMC3. Сравнение и области применения байесовского классификатора и логистической регрессии. Примеры задач бизнес-аналитики, сводящиеся к классификации алгоритмом “дерево решений”. Роль алгоритма в машинном обучении. Основной алгоритм “дерево решений”. Критерии ветвления. Алгоритм CART. Проблема переобучения. “Случайный лес”. Обзор существующих алгоритмов деревьев решений.
9.	Повышение точности классификаторов (ансамбли моделей)	Понятие неустойчивости решения, примеры. Обзор подходов к повышению точности (устойчивости) моделей. Понятие ансамбля моделей, примеры. Алгоритмы создания ансамблей моделей - бэккинг, буккинг (алгоритм AdaBoost.M1), случайный лес. Сравнение и области применения каждого метода.
10.	Оценка эффективности и	Необходимость оценки эффективности и точности моде-

	точности моделей	лей, примеры. Оценка ошибки моделей (кросс-валидация, бутстреп-оценка, исключение одного признака). Минимизация ошибок, приводящих к большим издержкам. Оптимизация модели по соотношению «издержки-прибыль» (Lift-диаграмма). Оптимизация бинарных моделей классификации (ROC-анализ). Обучение моделей при несбалансированных классах.
11.	Введение в машинное обучение (МО) и анализ данных (АД)	Бизнес-аналитика, решаемые задачи, большие данные. Технологии обработки больших данных. Введение в алгоритмы МО и АД.

Лабораторные занятия

№	Наименование раздела /темы дисциплины	Содержание
1	Тема 1.	Программная реализация генераторов случайных величин и построение гистограммы (язык Python)
2	Тема 2.	Расчет и отображение на гистограмме средней величины, СКО и доверительного интервала
3	Тема 3.	Программная реализация бутстрепа и построение гистограмм случайной величины и сроднённого значения
4	Тема 4.	Генерация двух случайных величин и построение двух гистограмм на одном графике
5	Тема 5.	Программная реализация гипотезы H_0 (отсутствует отличие между средними двух выборок)
6	Тема 6.	Программная реализация гипотезы H_1 (отличие между средними двух выборок существует)
7	Тема 7.	Расчет и визуализация ошибок 1-го и 2-го рода
8	Тема 8.	Сборка в единый модуль и выполнение программы проверки А/В гипотезы

4.3. Перечень учебно-методического обеспечения для самостоятельной работы студентов

1. Программное обеспечение и Интернет-ресурсы не предусмотрены.
2. Базы данных, информационно-справочные и поисковые системы не предусмотрены.
3. Жерон, Орельен Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем.: Пер. с англ. – СПб.: ООО «Альфа-книга», 2018. – 688 с.

5. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущего, рубежного и промежуточного контроля по дисциплине.

5.1. Паспорт фонда оценочных средств по дисциплине

Связь между формируемыми компетенциями и формами контроля их освоения представлена в следующей таблице:

Раздел	Темы занятий	Компетенция	Индикаторы освоения	Текущий контроль, неделя
Семестр 5				
1.	Тема 1. Введение в машинное обучение (МО) и анализ данных (АД)	ПК-5 ПК-14	3-ПК-5; У-ПК-5; В-ПК-5 3-ПК-14; У-ПК-14; В-ПК-14	УО 1,2
	Тема 2. Основы языка программирования Python			УО 2,3
	Тема 3. Сведения из математической статистики, необходимые для МО и АД			УО 3,4
	Тема 4. А/В гипотезы			УО 5
	Тема 5. Алгоритмы ассоциации			УО 6
Рубежный контроль		ПК-5 ПК-14	3-ПК-5; У-ПК-5; В-ПК-5 3-ПК-14; У-ПК-14; В-ПК-14	Тест 7
	Тема 6. Алгоритмы кластеризации	ПК-2.2	3-ПК-2.2 У-ПК-2.2; В-ПК-2.2	УО 8
2.	Тема 7. Линейная регрессия			УО 9
	Тема 8. Алгоритмы классификации (статистические методы, методы машинного обучения)			УО 10
	Тема 9. Повышение точности классификаторов (ансамбли моделей)			УО 11
	Тема 10. Оценка эффективности и точности моделей			УО 12
	Тема 11. Введение в машинное обучение (МО) и анализ данных (АД)			УО 13
Рубежный контроль		ПК-2.2	3-ПК-2.2 У-ПК-2.2; В-ПК-2.2	Тест 16
Промежуточная аттестация		ПК-5 ПК-14 ПК-2.2	3-ПК-5; У-ПК-5; В-ПК-5 3-ПК-14; У-ПК-14; В-ПК-14 3-ПК-2.2 У-ПК-2.2; В-ПК-2.2	экзамен

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

5.2.1. Оценочные средства для текущего контроля

5.2.1.1. Примерные темы для устного опроса

1. Типы данных. Массив библиотеки NumPy. Проверка типа данных.
2. Операторы if, for.
3. Построение графиков и гистограмм.
4. Случайные величины, функция распределения и функция плотности, примеры законов распределения из задач МО и АД - нормальный, биномиальный, Пуассона.
5. Параметры распределений: константы и/или случайные величины.
6. Доверительные интервалы.
7. Бутстреп и его возможности.

5.2.1.2. Примерные вопросы для самостоятельной работы

1. Метод бутстрепа для проверки статистической гипотезы о равенстве средних.
2. Алгоритм расчета ошибок 1-го и 2-го рода .
3. Алгоритм ассоциации «Априори».
4. Алгоритм кластеризации k-средних.
5. Алгоритм кластеризации G-средних.
6. Алгоритм кластеризации «Сеть Кохонера».
7. Алгоритм построения линейной регрессии.
8. Алгоритм классификации «Наивный Байес».
9. Алгоритм классификации методом опорных векторов (SVM).
10. Алгоритм построения дерева решений.
11. Алгоритмы обучения с подкреплением (МAB).

5.2.1.3. Примерные вопросы для тестов

1. Назначение метода бутстрепа.
2. Для каких задач предназначен алгоритм «Априори».
3. Отличие алгоритма «Априори» от иерархической ассоциации.
4. Сравнить алгоритмы k-средних и G-средних.
5. Алгоритм оптимизации весов с сети Кохонера.
6. Отличие линейной и логистической регрессии.
7. Понятие априорной и апостериорной вероятностей в байесовском классификаторе.
8. Понятия мягкого и жесткого зазоров в методе опорных векторов.
9. Критерий выбора узла для дальнейшего разбиения в деревьях решений.

10. Основное отличие алгоритма Томпсона (Thompson-Sampling) от ϵ -жадного и UCB.

5.2.3. Оценочные средства для промежуточной аттестации

5.2.3.1 Примерные вопросы к экзамену

1. Бизнес-аналитика, решаемые задачи, большие данные.
2. Технологии обработки больших данных.
3. Алгоритмы МО и АД.
4. Рекомендательные системы. Распознавание текста, лиц и речи.
5. Языки и системы программирования, используемые в машинном обучении.
6. Библиотеки для машинного обучения NumPy, SciPy, Matplotlib, Pandas.
7. Типы данных. Массив библиотеки NumPy. Проверка типа данных.
8. Операторы if, for.
9. Построение графиков и гистограмм.
10. Случайные величины, функция распределения и функция плотности, примеры законов распределения из задач МО и АД - нормальный, биномиальный, Пуассона.
11. Параметры распределений: константы и/или случайные величины.
12. Доверительные интервалы.
13. Бутстреп и его возможности.
14. Понятие статистической гипотезы, ошибки I и II рода.
15. Примеры задач бизнес-аналитики, сводящиеся к A/B гипотезам, понятия и примеры метрик.
16. Математическая (статистическая) постановка задачи, ошибки I и II рода.
17. Алгоритм аналитического решения для нормальных распределений.
18. Алгоритм численного (бутстреп) решения для произвольного (непараметрического) распределения и любой метрики.
19. Сравнение области применения каждого алгоритма.
20. Алгоритм ϵ -жадной (epsilon-greedy).
21. Алгоритм UCB, проблема в реализации и ее решение, сравнение с ϵ -жадным.

22. Алгоритм Томпсона (Thompson-Sampling), сравнение с ϵ -жадным и UCB.
23. Алгоритм, оптимизирующий выручку (прибыль).
24. Алгоритм Apriori.
25. Алгоритм иерархических ассоциативных правил.
26. Алгоритм последовательных шаблонов.
27. Сравнение и области применения каждого алгоритма.
28. Алгоритм k-средних (k-means), метрики –линейная и квадратичная, области применения каждой метрики.
29. Алгоритм G-mean, либо сеть Кохонена.
30. Сравнение и области применения каждого алгоритма.
31. Идея линейной регрессии, требования к исходным данным, оценка соответствия линейной модели данным (коэффициент корреляции, t- и F-критерии).
32. Построение линейной регрессии средствами Python (scipy).
33. Многомерная линейная регрессия, понижение размерности.
34. Ограничения и область применения линейной регрессии в задачах бизнес аналитики.
35. Алгоритм метода опорных векторов (SVM).
36. Логистическая регрессия.
37. Сравнение и области применения каждого алгоритма.
38. Алгоритм байесовского классификатора.
39. Вводные понятия решения этой задачи пакетом PyMC3.
40. Сравнение и области применения байесовского классификатора и логистической регрессии.
41. Роль алгоритма в машинном обучении.
42. Алгоритм “дерево решений”.
43. Критерии ветвления.
44. Алгоритм CART.

45. Проблема переобучения.
46. “Случайный лес”.
47. Существующие алгоритмы деревьев решений.
48. Понятие неустойчивости решения, примеры.
49. Подходы к повышению точности (устойчивости) моделей.
50. Понятие ансамбля моделей, примеры.
51. Алгоритмы создания ансамблей моделей - бэккинг, буккинг, случайный лес.
52. Сравнение и области применения каждого метода.
53. Необходимость оценки эффективности и точности моделей, примеры.
54. Оценка ошибки моделей.
55. Минимизация ошибок, приводящих к большим издержкам.
56. Оптимизация модели по соотношению «издержки-прибыль» (Lift-диаграмма).
57. Оптимизация бинарных моделей классификации (ROC-анализ).
58. Обучение моделей при несбалансированных классах.

5.3. Шкалы оценки образовательных достижений

Рейтинговая оценка знаний является интегральным показателем качества теоретических и практических знаний и навыков студентов по дисциплине и складывается из оценок, полученных в ходе текущего контроля и промежуточной аттестации.

Результаты текущего контроля и промежуточной аттестации подводятся по шкале балльно-рейтинговой системы.

Шкала каждого контрольного мероприятия лежит в пределах от 0 до установленного максимального балла включительно. Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет собой сумму баллов, заработанных студентом при выполнении заданий в рамках текущего и промежуточного контроля.

Итоговая оценка выставляется в соответствии со следующей шкалой:

Сумма баллов	Оценка по 4-ех балльной шкале	Оценка ECTS	Требования к уровню освоению учебной дисциплины
90-100	5 – «отлично»	A	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил

			программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, использует в ответе материал монографической литературы.
85-89	4 – «хорошо»	В	Оценка «хорошо» выставляется студенту, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос.
75-84		С	
70-74		Д	
65-69	3 – «удовлетворительно»	Е	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала.
60-64			
Ниже 60	2 – «неудовлетворительно»	Ф	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям: Учебное пособие. 2-е изд., испр. – СПб.: Питер, 2013. – 704 с.
2. Барсегян А.А. Анализ данных и процессов: учеб. пособие/ 3-е изд. - СПб.: БХВ-Петербург, 2009. – 512 с.
3. Брюс П. Практическая статистика для специалистов Data Science: Пер. с англ. / П. Брюс, Э. Брюс. – СПб.: БХВ – Петербург, 2018. – 304 с.
4. Грас Дж. Data Science. Наука о данных с нуля: Пер. с англ. – СПб.: БХВ -Петербург, 2017. - 336 с.
5. Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер. 2018. – 576 с.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. . Программное обеспечение и Интернет-ресурсы не предусмотрены.

2. Базы данных, информационно-справочные и поисковые системы не предусмотрены.
3. Жерон, Орельен Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем.: Пер. с англ. – СПб.: ООО «Альфа-книга», 2018. – 688 с.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ и интернет- ресурсы

1. Программная среда «Anaconda», (Open Source).
2. Доступ в Интернет для обращения к документации по библиотекам, входящим в состав среды «Anaconda».

7 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Лекционные занятия проходят в аудитории, оснащенной компьютером и проектором. Каждая лекция сопровождается презентацией, содержащей теоретический материал.

Изучение дисциплины проводится в лабораториях кафедры «Цифровых технологий». Лабораторные работы проводятся с использованием ресурсов компьютерных классов, позволяющих работать в средах MS Windows, при использовании программных средств

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При чтении лекционного материала используется электронное сопровождение курса: справочно-иллюстративный материал воспроизводится и озвучивается в аудитории с использованием проектора и переносного компьютера в реальном времени. На сайте кафедры также находится методический и справочный материал, необходимый для проведения практических работ по курсу.

9. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ОРГАНИЗАЦИИ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ

Примерным учебным планом на изучение дисциплины отводится один семестр. В конце предусмотрен экзамен.

При преподавании дисциплины методически целесообразно в каждом разделе курса выделить наиболее важные моменты и акцентировать на них внимание обучающихся.

Предлагается:

- В первом разделе заострить внимание на понятиях машинное обучение (МО) и анализ данных (АД). Введение в алгоритмы МО и АД.
- Во втором разделе рассмотреть язык программирования Python используемый в машинном обучении. Изучить операторы if, for, а также построение графиков и гистограмм.

- В третьем разделе изучить сведения из математической статистики: случайные величины, функции распределения и функции плотности, примеры законов распределения из задач МО и АД. Доверительные интервалы. Бутстреп и его возможности. Понятие статистической гипотезы, ошибки I и II рода.
- В четвертом разделе особое внимание уделить вопросам A/B гипотезы. Рассмотреть алгоритмы аналитического и численного решения и сравнить области применения каждого алгоритма.
- В пятом разделе изложить общие принципы алгоритма MAB. Рассмотреть алгоритмы ϵ -жадной, UCB, Томпсона, оптимизирующий выручку.
- В разделе шесть рассмотреть вопросы, связанные с алгоритмами ассоциации. Изучить алгоритмы Apriori, иерархических ассоциативных правил, последовательных шаблонов и сравнить области применения каждого из алгоритмов.
- В седьмом разделе большое внимание следует уделить алгоритмам кластеризации. Рассмотреть алгоритмы k-средних, G-mean, сеть Кохонена. А также сравнить их и определить области применения каждого из алгоритмов.
- В разделе восемь изучить сведения о линейной регрессии. Идея линейной регрессии, требования к исходным данным, оценка соответствия линейной модели данным. Построение линейной регрессии. Рассмотреть ограничения и области применения линейной регрессии, в задачах бизнес аналитики.
- В девятом разделе необходимо рассмотреть алгоритмы классификации. Статистические методы и методы машинного обучения. Изучить алгоритм байесовского классификатора их роль, понятия решения этой задачи, сравнение и области применения. Уделить внимание на алгоритм CART и выполнить обзор существующих алгоритмов деревьев решений.
- В разделе десять изучить подходы к повышению точности классификатора. Рассмотреть понятие ансамбля модели и алгоритмы их создания.
- В одиннадцатом разделе необходимо оценить эффективность и точность моделей. Изучить ошибки моделей. Оптимизировать бинарные модели и модели по соотношению «издержки-прибыль».

Программа составлена в соответствии с требованиями ОС ВО НИЯУ МИФИ к обязательному минимуму содержания основной образовательной программы по направлению подготовки 09.03.02 Информационные системы и технологии.

Программу составил:

Рецензент: