

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Гриб Владислав Валерьевич  
Должность: Ректор  
Дата подписания: 03.11.2023 16:59:02  
Уникальный программный ключ:  
637517d24e103c3db032acf37e0b6498ec1c5bb2f5ab80c39ebfad7f47095447



**Образовательное частное учреждение высшего образования**  
**«МОСКОВСКИЙ УНИВЕРСИТЕТ ИМЕНИ А.С. ГРИБОЕДОВА»**  
(ИМПЭ им. А.С. Грибоедова)

**Институт международной экономики, лидерства и менеджмента**

**УТВЕРЖДАЮ**  
Директор института  
международной экономики,  
лидерства и менеджмента  
\_\_\_\_\_ А.А. Панарин  
«28» сентября 2023 г.

**Рабочая программа дисциплины**  
**ИЗБРАННЫЕ ВОПРОСЫ АНАЛИЗА ДАННЫХ**

**Направление подготовки 09.03.03 Прикладная информатика**  
**(уровень бакалавриат)**

**Направленность (профиль):**  
**«Анализ данных»**

**Форма обучения: очная**

**Москва**

Рабочая программа дисциплины «Избранные вопросы анализа данных». Направление подготовки 09.03.03 Прикладная информатика, направленность (профиль): «Анализ данных» / А.А. Шестемиров – М.: ИМПЭ им. А.С. Грибоедова. – 22 с.

Рабочая программа дисциплины составлена на основании федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 09.03.03 Прикладная информатика, утвержденного приказом Министерства образования и науки Российской Федерации от «19» сентября 2017 № 922 (с изменениями и дополнениями) и Профессионального стандарта «Программист», утвержденного приказом Министерства труда и социальной защиты Российской Федерации от 18 ноября 2013 г. № 679н (зарегистрирован Министерством юстиции Российской Федерации 18 декабря 2013 г., регистрационный № 30635), с изменением, внесенным приказом Министерства труда и социальной защиты Российской Федерации от 12 декабря 2016 г. № 727н (зарегистрирован Министерством юстиции Российской Федерации 13 января 2017 г., регистрационный № 45230), Профессионального стандарта «Специалист по информационным системам», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 18 ноября 2014 г. № 896н (зарегистрирован Министерством юстиции Российской Федерации 24 декабря 2014 г., регистрационный № 35361), с изменением, внесенным приказом Министерства труда и социальной защиты Российской Федерации от 12 декабря 2016 г. № 727н (зарегистрирован Министерством юстиции Российской Федерации 13 января 2017 г., регистрационный № 45230)

Разработчики:

К.э.н. А.А. Шестемиров

Ответственный рецензент:

Назарова Н.А., к.э.н., доцент, заместитель руководителя департамента налогов и налогового администрирования Финансового университета при Правительстве Российской Федерации

*(Ф.И.О., уч. степень, уч. звание, должность)*

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры инновационного менеджмента и предпринимательства 15.09.2023г., протокол №2

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_ /к.э.н. А.А. Шестемиров/

(подпись)

Согласовано от Библиотеки \_\_\_\_\_ /О.Е. Степкина/

(подпись)

## РАЗДЕЛ 1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Избранные вопросы анализа данных» является овладение студентами навыками решения задач по разработке алгоритмов анализа и обработки данных.

Задачи, решаемые в ходе освоения программы дисциплины:

- изучение принципов и методов обработки данных;
- ознакомление с современными средствами обработки и анализа данных;
- овладение первоначальными навыками построения моделей для решения прикладных задач.

## РАЗДЕЛ 2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

### ПРОФЕССИОНАЛЬНЫЕ КОМПЕТЕНЦИИ

Код, наименование профессиональных компетенций	Трудовые функции (код, наименование)/уровень (подуровень) квалификации	Индикаторы достижения компетенции (для планирования результатов обучения по элементам образовательной программы и соответствующих оценочных средств)
<b>ОТФ (код, наименование) / Профессиональный стандарт (код, наименование)</b>		
<b>Тип(ы) задач(и) профессиональной деятельности</b>		
D. Разработка требований и проектирование программного обеспечения/ 06.001 Программист		
Проектный		
ПК-1. Способность проектировать прикладное программное обеспечение	D/03.6 Проектирование программного обеспечения	<p>ИПК-1.1 Знать принципы построения архитектуры программного обеспечения и виды архитектуры программного обеспечения; типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения; методы и средства проектирования программного обеспечения; методы и средства проектирования баз данных; методы и средства проектирования программных интерфейсов</p> <p>ИПК-1.2 Уметь использовать существующие типовые решения и шаблоны проектирования программного обеспечения; применять методы и средства проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов; осуществлять коммуникации с заинтересованными сторонами</p> <p>ИПК-1.3 Владеть разработкой, изменением и согласованием архитектуры программного обеспечения с системным аналитиком и архитектором программного обеспечения; системой проектирования структур данных; системой проектирования баз данных; системой проектирования программных интерфейсов; оценкой и согласованием сроков выполнения поставленных задач</p>
С Выполнение работ и управление работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы / 06.015 Профессиональный стандарт «Специалист по информационным системам»		
производственно-технологический , научно-исследовательский		
ПК-2. Способность осуществлять инженерно-технологическую поддержку планирования управления требованиями по созданию (модификации) и сопровождению ИС	С/10.6 Инженерно-технологическая поддержка планирования управления требованиями	ИПК-2.1. Знать инструменты и методы управления требованиями; предметной областью автоматизации; возможностями ИС; источниками информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; инструменты и методы моделирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта: документирование требований, систему анализа продукта, модерируемых совещаний; управление качеством: контрольные списки, верификация, валидация (приемо-сдаточные испытания); управление коммуникациями в проекте: базовые навыки управления (в том числе

		<p>проведение презентаций, проведение переговоров, публичные выступления); культуру речи; правила деловой переписки.</p> <p>ИПК-2.2. Умеет анализировать входные данные; планировать работы.</p> <p>ИПК-2.3. Владеть системным подходом для выбора технологии управления требованиями; представлениями исходных данных для разработки плана управления требованиями; системой согласования в части инженерно-технологического обеспечения плана управления требованиями с заинтересованными сторонами</p>
<p>ПК-3. Способность применять системный подход для составления технико-экономического обоснования проектных решений и технического задания на разработку информационной системы</p>	<p>С/01.6 Определение первоначальных требований заказчика к ИС и возможности их реализации в ИС на этапе предконтрактных работ;</p>	<p>ИПК3.1. Знает возможности типовой ИС; предметную область автоматизации; методы выявления требований; технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии; технологии подготовки и проведения презентаций; архитектуру, устройство и функционирование вычислительных систем; коммуникационное оборудование; сетевые протоколы; основы современных операционных систем; основы современных систем управления базами данных; устройство и функционирование современных ИС; современные стандарты информационного взаимодействия систем; программные средства и платформы инфраструктуры информационных технологий организаций; современные подходы и стандарты автоматизации организации (например, CRM, MRP, ERP..., ITIL, ITSM); основы теории систем и системного анализа; методики описания и моделирования бизнес-процессов, средства моделирования бизнес-процессов; системы классификации и кодирования информации, в том числе присвоение кодов документам и элементам справочников; отраслевая нормативная техническая документация; источники информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; формирование и механизмы рыночных процессов организации; основы менеджмента, в том числе менеджмента качества; основы бухгалтерского учета и отчетности организаций; основы налогового законодательства Российской Федерации; основы управленческого учета; основы финансового учета и бюджетирования; основы международных стандартов финансовой отчетности (МСФО); основы управления торговлей, поставками и запасами; основы организации производства; основы управления взаимоотношениями с клиентами и заказчиками (CRM); основы управления персоналом, включая вопросы оплаты труда; основы теории управления; современные инструменты и методы управления организацией, в том числе методы планирования деятельности, распределения поручений, контроля исполнения, принятия решений; методология ведения документооборота в организациях; инструменты и методы определения финансовых и производственных показателей деятельности организаций; основы организационной диагностики; инструменты и методы моделирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта документирование требований, анализ продукта, модерлируемые совещания; культура речи; правила деловой переписки.</p> <p>ИПК-3.2 Умеет проводить переговоры; проводить презентации; подготавливать протоколы мероприятий.</p> <p>ИПК 3.3. Владеет основами выявления первоначальных требований заказчика к ИС; системой информирования заказчика о возможностях типовой ИС и вариантах ее модификации;</p>

		определениями возможностей достижения соответствия ИС первоначальным требованиям заказчика; основами составления протоколов переговоров с заказчиком
--	--	--

### РАЗДЕЛ 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Избранные вопросы анализа данных» изучается в седьмом семестре, относится к Б1.В. Части учебного плана, формируемой участниками образовательных отношений, Блока Б.1 «Дисциплины (модули)».

Общая трудоемкость дисциплины составляет 3 з.е.

Знания, умения, навыки, опыт практической деятельности, приобретенные при освоении настоящей дисциплины, необходимы для успешного освоения следующих дисциплин: «Интеллектуальные информационные системы», «Системы анализа данных», «Компьютерные экспертные системы».

### РАЗДЕЛ 4. ОБЪЕМ (ТРУДОЕМКОСТЬ) ДИСЦИПЛИНЫ

(ОБЩАЯ, ПО ВИДАМ УЧЕБНОЙ РАБОТЫ, ВИДАМ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ)

#### Трудоемкость дисциплины и виды учебной работы (очная форма обучения)

З.е.	Всего часов	Контактная работа				Часы СР на подготовку кур.раб.	Иная СР	Контроль
		Занятия лекционного типа	Занятия семинарского типа		Контактная работа по курсовой работе			
			Лабораторные	Практические/Семинарские				
7 семестр								
3	108	20	32	-	-	-	54	2 зачет
Всего по дисциплине								
3	108	20	32	-	-	-	54	2

### СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

№ п/п	Наименование разделов и тем дисциплины	Содержание темы
<b>Раздел №1 «Введение в машинное обучение»</b>		
1	Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения	Изучаемые вопросы: 1. Понятие машинного обучения. 2. Этапы решения задач методами машинного обучения. 3. Постановка задачи машинного обучения с учителем. 4. Сложности, связанные с применением методов машинного обучения. Вопросы для самостоятельного изучения: 1. Примеры задач машинного обучения.
2	Тема 1.2: Предварительная обработка данных. Проектирование признаков	Изучаемые вопросы: 1. Особенности сбора данных для решения задач машинного обучения. 2. Подготовка данных к моделированию. Категориальные признаки. 3. Обработка отсутствующих значений. 4. Нормализация признаков. 5. Визуализация. 6. Выбор признаков. Вопросы для самостоятельного изучения: 1. Усовершенствованное проектирование признаков.
<b>Раздел №2 «Модели и алгоритмы машинного обучения»</b>		

№ п/п	Наименование разделов и тем дисциплины	Содержание темы
3	Тема 2.1 Алгоритмы восстановления пропусков	Изучаемые вопросы: 1. Постановка задачи восстановления пропусков. 2. Алгоритм ZET. Вопросы для самостоятельного изучения: 1. Анализ данных с пропусками на основе функции правдоподобия
4	Тема 2.2 Задача понижения размерности.	Изучаемые вопросы: 1. Понятие признакового пространства. 2. Постановка задачи понижения размерности признакового пространства. 3. Методы понижения размерности. 4. Метод главных компонент. Вопросы для самостоятельного изучения: 1. Метод факторного анализа.
5	Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии	Изучаемые вопросы: 1. Наивный байесовский классификатор. 2. Метрические методы классификации и регрессии. Вопросы для самостоятельного изучения: 1. Преимущества и проблемы применения метода k-ближайших соседей.
6	Тема 2.4 Логические методы классификации. Деревья решений	Изучаемые вопросы: 1. Основные понятия. 2. Алгоритм построения решающих деревьев ID3. 3. Обработка пропусков в решающих деревьях. 4. Переусложнение структуры дерева. Прунинг. 5. Деревья для регрессии. Вопросы для самостоятельного изучения: 1. Алгоритм CART.
7	Тема 2.5 Ансамбли моделей	Изучаемые вопросы: 1. Баггинг и случайные леса. 2. Усиление. Обучение усиленных правил. Вопросы для самостоятельного изучения: 1. Карта ансамблевого ландшафта.
8	Тема 2.6 Задача кластеризации	Изучаемые вопросы: 1. Общая постановка задача кластеризации. 2. Обзор алгоритмов кластеризации. 3. Алгоритм ISODATA. 4. Плотностные алгоритмы DBSCAN, DENCLUE. 5. Валидация результатов кластеризации. Вопросы для самостоятельного изучения: 1. Алгоритмы кластеризации `больших` данных.
9	Тема 2.7 Обучение с подкреплением	Изучаемые вопросы: 1. Обучение с подкреплением. 2. Элементы обучения с подкреплением. 3. Оценочная обратная связь. 4. Динамическое программирование. 5. Методы Монте-Карло. Вопросы для самостоятельного изучения: 1. Задача об одnorуком бандите. 2. Обучение на основе временных различий

**Перечень разделов (модулей), тем дисциплины и распределение учебного времени по разделам/темам дисциплины, видам учебных занятий (в т.ч. контактной работы), видам текущего контроля очная форма обучения**

Разделы / Темы	Контактная работа	Часы СР на	Иная СР	Кон- троль	Всего часов
----------------	-------------------	---------------	------------	---------------	----------------

	Занятия лекционного типа	Занятия семинарского типа		Контактная работа по курсу	подготовка к курсу	Б	
		Лаб. р	Практик./сем.				
<b>7 семестр</b>							
<b>Раздел 1. «Введение в машинное обучение»</b>	<b>4</b>	<b>4</b>				<b>8</b>	<b>16</b>
Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения	2					4	6
Тема 1.2 Предварительная обработка данных. Проектирование признаков	2	4				4	10
<b>Раздел №2 «Модели и алгоритмы машинного обучения»</b>	<b>16</b>	<b>28</b>				<b>46</b>	<b>90</b>
Тема 2.1 Алгоритмы восстановления пропусков	2	4				6	12
Тема 2.2 Задача понижения размерности	2	4				4	10
Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии	2	4				6	12
Тема 2.4 Логические методы классификации. Деревья решений	2	4				10	16
Тема 2.5 Ансамбли моделей	2	4				7	13
Тема 2.6 Задача кластеризации	2	4				4	10
Тема 2.7 Обучение с подкреплением	4	4				6	16
Зачет							2
<b>Итого за 7 семестр</b>	<b>20</b>	<b>32</b>				<b>54</b>	<b>108</b>

## ЗАНЯТИЯ СЕМИНАРСКОГО ТИПА для очной формы обучения

### *Семинарские занятия*

Общие рекомендации по подготовке к семинарским занятиям. При подготовке к работе во время проведения занятий семинарского типа следует обратить внимание на следующие моменты: на процесс предварительной подготовки, на работу во время занятия, обработку полученных результатов, исправление полученных замечаний. Предварительная подготовка к учебному занятию семинарского типа заключается в изучении теоретического материала в отведенное для самостоятельной работы время, ознакомление с инструктивными материалами с целью осознания задач занятия. Работа во время проведения занятия семинарского типа включает несколько моментов: а) консультирование обучающихся преподавателями с целью предоставления исчерпывающей информации, необходимой для самостоятельного выполнения предложенных преподавателем задач, б) самостоятельное выполнение заданий согласно обозначенной учебной программой тематики.

### 7 семестр

#### Раздел №1 «Введение в машинное обучение»

**Лабораторная работа 1. Преобразование признаков. Конструирование и отбор признаков (4 ч).**

**Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Выделить признаки с годом, месяцем, неделями, днем;

3. Обработать текстовые признаки.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постова-лов, Е. В. Чимитова. — Новосибирск : Новосибирский государственный технический университет, 2011. — 888 с. — ISBN 978-5-7782-1590-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/47719.html>

**Раздел №2 «Модели и алгоритмы машинного обучения»**

**Лабораторная работа 2. Задача восстановления пропусков (4 ч).**

**Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Восстановить пропущенные значения.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постова-лов, Е. В. Чимитова. — Новосибирск : Новосибирский государственный технический университет, 2011. — 888 с. — ISBN 978-5-7782-1590-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/47719.html>

**Лабораторная работа 3. Задача понижения размерности (4 ч).**

**Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Реализовать метод главных компонент.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

**Лабораторная работа 4. Задача классификации: вероятностные модели (4 ч).**

**Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Реализовать модель логистической регрессии и наивного байесовского классификатора.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>



Дополнительная

2. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

#### **Лабораторная работа 5. Задача классификации: деревья решений (4 ч).**

##### **Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Реализовать модель деревьев решений.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

#### **Лабораторная работа 6. Задача классификации: случайный лес (4 ч).**

##### **Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Реализовать модель случайного леса.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

#### **Лабораторная работа 7. Задача кластеризации (4 ч).**

##### **Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Реализовать алгоритмы иерархической и неиерархической кластеризации.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

#### **Лабораторная работа 8. Задача кластеризации больших данных (4 ч).**

##### **Порядок выполнения работы:**

1. Изучить выданный файл с исходными данными – матрица «объектов-признаков» с метками;
2. Реализовать алгоритм DBSCAN.

Литература:

Основная

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>

Дополнительная

2. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

## РАЗДЕЛ 5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

### Интерактивные образовательные технологии, используемые на аудиторных лабораторных занятиях

#### Очная форма обучения

Наименование разделов, тем	Используемые образовательные технологии	Часы
<b>Раздел №1 «Введение в машинное обучение»</b> Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков	Обсуждение решений профессионально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2
<b>Раздел №2 «Модели и алгоритмы машинного обучения»</b> Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с подкреплением	Обсуждение решений профессионально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2

## РАЗДЕЛ 6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

### Самостоятельная работа

Наименование разделов/тем	Вопросы, выносимые на самостоятельное изучение
<b>Раздел №1 «Введение в машинное обучение»</b> Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков	Примеры задач машинного обучения. Усовершенствованное проектирование признаков
<b>Раздел №2 «Модели и алгоритмы машинного обучения»</b> Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации	Анализ данных с пропусками на основе функции правдоподобия. Метод факторного анализа. Преимущества и проблемы применения метода k-ближайших соседей. Алгоритм CART. Карта ансамблевого ландшафта. Алгоритмы кластеризации «больших» данных. Задача об одnorуком бандите.

Наименование разделов/тем	Вопросы, выносимые на самостоятельное изучение
Тема 2.7 Обучение с подкреплением	Обучение на основе временных различий.

### 6.1. Примерные задания для самостоятельной работы

#### 1. Задача на построение модели классификации

Дано: файл с исходными данными – матрица «объектов-признаков» с метками (для моделей типа «обучение с учителем»)

Необходимо:

выполнить предварительную обработку данных и проектирование признаков, при необходимости;

выполнить обучение модели классификации методом  $k$ -ближайших соседей при трех различных значения параметра  $k$ ;

выполнить оценку точности модели методом кросс-валидации по 5 и 10 блокам, проанализировать результаты и сделать выводы о качестве построенной модели.

#### 2. Задача на построение модели регрессии

Дано: файл с исходными данными – матрица «объектов-признаков» с метками (для моделей типа «обучение с учителем»)

Необходимо:

выполнить предварительную обработку данных и проектирование признаков, при необходимости;

выполнить обучение линейной модели регрессии с использованием  $L_2$ -регуляризации (гребневая регрессия) с тремя различными параметрами  $\alpha$ ;

выполнить прогноз значений меток для тестовой (отложенной) части матрицы «объектов-признаков»;

выполнить оценку точности модели отложенных данных, проанализировать результаты и сделать выводы о качестве построенной модели.

## РАЗДЕЛ 7. ОЦЕНОЧНЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЕ (ФОНД ОЦЕНОЧНЫХ СРЕДСТВ) ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

### 7.1. Форма промежуточной аттестации обучающегося по учебной дисциплине.

Контрольным мероприятием промежуточной аттестации обучающихся по учебной дисциплине «Избранные вопросы анализа данных» в 7 семестре является зачет, который проводится в устной форме.

Таблица 7.1

### ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ В СООТНОШЕНИИ С ОЦЕНОЧНЫМИ СРЕДСТВАМИ

Планируемые результаты, характеризующие этапы формирования компетенции	Содержание учебного материала	Примеры контрольных вопросов и заданий для оценки знаний, умений, владений	Методы/ средства контроля
ПК-1 Способность проектировать прикладное программное обеспечение			
ИПК-1.1 Знать принципы построения архитектуры программного обеспечения и виды архитектуры программного обеспечения; типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения; методы и средства проектирования программного обеспечения; методы и средства проектирования баз данных; методы и средства проектирования программных	Раздел 1. «Введение в машинное обучение» Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков	Понятие машинного обучения. Типы машинного обучения и виды задач, решаемых в рамках данного направления. Этапы решения задач методами машинного обучения. Межотраслевой стандарт решения задач интеллектуального анализа данных.	Устный контроль/ опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ защиты проектов; применение

интерфейсов	<p>Раздел №2 «Модели и алгоритмы машинного обучения»</p> <p>Тема 2.1 Алгоритмы восстановления пропусков</p> <p>Тема 2.2 Задача понижения размерности</p> <p>Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии</p> <p>Тема 2.4 Логические методы классификации. Деревья решений</p> <p>Тема 2.5 Ансамбли моделей</p> <p>Тема 2.6 Задача кластеризации</p> <p>Тема 2.7 Обучение с подкреплением</p>	<p>Постановка задачи машинного обучения с учителем. Типы признаков.</p> <p>Понятие эмпирического риска и сведение задачи обучения к задаче оптимизации.</p> <p>Примеры задач машинного обучения в различных прикладных областях (медицина, финансовая сфера и т.д.)</p>	<p>теоретических знаний при анализе (разборе) конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ использования теоретических знаний в процессе решения кейсов, в ходе деловых игр; письменный контроль, анализ содержания эссе; тестирование (выполнение тестовых заданий)</p>
ИПК-1.2 Уметь использовать существующие типовые решения и шаблоны проектирования программного обеспечения; применять методы и средства проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов; осуществлять коммуникации с заинтересованными сторонами	<p>Раздел 1. «Введение в машинное обучение»</p> <p>Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения</p> <p>Тема 1.2 Предварительная обработка данных. Проектирование признаков</p> <p>Раздел №2 «Модели и алгоритмы машинного обучения»</p> <p>Тема 2.1 Алгоритмы восстановления пропусков</p> <p>Тема 2.2 Задача понижения размерности</p> <p>Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии</p> <p>Тема 2.4 Логические методы классификации. Деревья решений</p> <p>Тема 2.5 Ансамбли моделей</p> <p>Тема 2.6 Задача кластеризации</p> <p>Тема 2.7 Обучение с подкреплением</p>	<p>Метрические методы классификации и регрессии. Метрические классификаторы в задачах восстановления регрессии. Переход к непараметрической регрессии.</p> <p>Формальная постановка задачи линейной регрессии и классификации. Наиболее часто используемые функции для аппроксимации пороговой функции. Градиентный метод численной минимизации в задачах линейной классификации. Метод стохастического градиента.</p> <p>Линейные модели регрессии. Применение линейной регрессии к нелинейным отношениям между переменными.</p>	<p>Анализ проявленных умений при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ</p>
ИПК-1.3 Владеть разработкой, изменением и согласованием архитектуры программного обеспечения с системным аналитиком и архитектором программного обеспечения; системой проектирования структур данных; системой проектирования баз данных; системой проектирования программных интерфейсов; оценкой и согласованием сроков выполнения поставленных задач	<p>Раздел 1. «Введение в машинное обучение»</p> <p>Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения</p> <p>Тема 1.2 Предварительная обработка данных. Проектирование признаков</p> <p>Раздел №2 «Модели и алгоритмы машинного обучения»</p> <p>Тема 2.1 Алгоритмы восстановления пропусков</p> <p>Тема 2.2 Задача понижения</p>	<p>Композиции алгоритмов. Простое голосование базовых классификаторов. Методы повышения различности классификаторов.</p> <p>Случайный лес.</p> <p>Градиентный бустинг. Алгоритм AdaBoost.</p> <p>Искусственные нейронные сети в решении задач машинного обучения. Линейная модель нейрона.</p> <p>Нейронная реализация логических функций.</p>	<p>Анализ проявленных навыков при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий);</p>

	<p>размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с подкреплением</p>		<p>анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ</p>
<p>ПК-2. Способность осуществлять инженерно-технологическую поддержку планирования управления требованиями по со-зданию (модификации) и сопровождению ИС</p>			
<p>ИПК-2.1. Знать инструменты и методы управления требованиями; предметной областью автоматизации; возможностями ИС; источниками информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; инструменты и методы моделирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта: документирование требований, систему анализа продукта, модерлируемых совещаний; управление качеством: контрольные списки, верификация, валидация (приемо-сдаточные испытания); управление коммуникациями в проекте: базовые навыки управления (в том числе проведение презентаций, проведение переговоров, публичные выступления); культуру речи; правила деловой переписки.</p>	<p>Раздел 1. «Введение в машинное обучение» Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков Раздел №2 «Модели и алгоритмы машинного обучения» Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с подкреплением</p>	<p>Особенности сбора данных для решения задач машинного обучения: число признаков, количество и качество обучающей выборки. Предварительная обработка данных. Обработка категориальных признаков. Предварительная обработка данных. Обработка отсутствующих значений. Предварительная обработка данных. Нормализация признаков. Предварительная обработка данных. Стратегии выбора признаков. Визуализация взаимосвязей между признаками и целевой переменной. Основы проектирования признаков.</p>	<p>Устный контроль/ опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ защиты проектов; применение теоретических знаний при анализе (разборе) конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ использования теоретических знаний в процессе решения кейсов, в ходе деловых игр; письменный контроль, анализ содержания эссе; тестирование (выполнение тестовых заданий)</p>
<p>ИПК-2.2. Умеет анализировать входные данные; планировать работы.</p>	<p>Раздел 1. «Введение в машинное обучение» Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков Раздел №2 «Модели и алгоритмы машинного обучения» Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы класси-</p>	<p>Усовершенствованное проектирование признаков. Оптимизация модели путем подбора параметров. Алгоритм сеточного поиска. Байесовская классификация. Гауссов и полиномиальный наивный байесовский классификаторы. Метрические методы классификации и регрессии. Метод k-ближайших соседей (kNN). Метрические методы классификации и регрессии. Метод k-ближайших соседей (kNN). Метод окна Парзена. Метрические методы классификации и регрессии.</p>	<p>Анализ проявленных умений при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных</p>

	<p>фикации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с подкреплением</p>	<p>Метрические классификаторы в задачах восстановления регрессии. Переход к непараметрической регрессии.</p>	<p>проектов; опрос на семинарских занятиях, зачете, анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ</p>
<p>ИПК-2.3. Владеть системным подходом для выбора технологии управления требованиями; представлениями исходных данных для разработки плана управления требованиями; системой согласования в части инженерно-технологического обеспечения плана управления требованиями с заинтересованными сторонами</p>	<p>Раздел 1. «Введение в машинное обучение» Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков Раздел №2 «Модели и алгоритмы машинного обучения» Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с подкреплением</p>	<p>Формальная постановка задачи линейной регрессии и классификации. Наиболее часто используемые функции для аппроксимации пороговой функции. Градиентный метод численной минимизации в задачах линейной классификации. Метод стохастического градиента. Линейные модели регрессии. Применение линейной регрессии к нелинейным отношениям между переменными. Проблема мультиколлинеарности. Разновидности линейных моделей и применение регуляризаторов. Гребневая регрессия (Ridge, L2-регуляризация). LASSO (L1-регуляризация).</p>	<p>Анализ проявленных навыков при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ</p>
<p>ПК-3. Способность применять системный подход для составления технико-экономического обоснования проектных решений и технического задания на разработку информационной системы</p>			
<p>ИПК3.1. Знает возможности типовой ИС; предметную область автоматизации; методы выявления требований; технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии; технологии подготовки и проведения презентаций; архитектуру, устройство и функционирование вычислительных систем; коммуникационное оборудование; сетевые протоколы; основы современных операционных систем; основы современных систем управления базами данных; устройство и функционирование современных ИС; современные стандарты информационного взаимодействия систем; программные средства и платформы инфраструктуры информационных технологий организаций; современные подходы и стандарты автоматизации</p>	<p>Раздел 1. «Введение в машинное обучение» Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков Раздел №2 «Модели и алгоритмы машинного обучения» Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесов-</p>	<p>Признаки, извлекаемые из текстов и изображений. Проблема переобучения. Методы оценки модели. Оценка на отложенных данных. Методы оценки модели. Кросс-валидация по к-блокам. Оценка моделей бинарной классификации. Таблицы сопряженности. ROC-кривая. Оценка моделей многоклассовой классификации. Таблицы сопряженности. ROC-кривые. 7. Оценка моделей регрес-</p>	<p>Устный контроль/ опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ защиты проектов; применение теоретических знаний при анализе (разборе) конкретных практико-ориентированных ситуаций и профессионально-прикладных задач,</p>

<p>ции организации (например, CRM, MRP, ERP..., ITIL, ITSM); основы теории систем и системного анализа; методики описания и моделирования бизнес-процессов, средства моделирования бизнес-процессов; системы классификации и кодирования информации, в том числе присвоение кодов документам и элементам справочников; отраслевая нормативная техническая документация; источники информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; формирование и механизмы рыночных процессов организации; основы менеджмента, в том числе менеджмента качества; основы бухгалтерского учета и отчетности организаций; основы налогового законодательства Российской Федерации; основы управленческого учета; основы финансового учета и бюджетирования; основы международных стандартов финансовой отчетности (МСФО); основы управления торговлей, поставками и запасами; основы организации производства; основы управления взаимоотношениями с клиентами и заказчиками (CRM); основы управления персоналом, включая вопросы оплаты труда; основы теории управления; современные инструменты и методы управления организацией, в том числе методы планирования деятельности, распределения поручений, контроля исполнения, принятия решений; методология ведения документооборота в организациях; инструменты и методы определения финансовых и производственных показателей деятельности организаций; основы организационной диагностики; инструменты и методы моделирования бизнес-процессов организации; основы ре-инжиниринга бизнес-процессов организации; управление содержанием проекта документирование требований, анализ продукта, модерлируемые совещания; культура речи; правила деловой переписки.</p>	<p>ский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с подкреплением</p>	<p>сии.</p>	<p>анализ использования теоретических знаний в процессе решения кейсов, в ходе деловых игр; письменный контроль, анализ содержания эссе; тестирование (выполнение тестовых заданий)</p>
<p>ИПК-3.2 Умеет проводить переговоры; проводить презентации; подготавливать протоколы мероприятий.</p>	<p>Раздел 1. «Введение в машинное обучение» Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения Тема 1.2 Предварительная обработка данных. Проектирование признаков Раздел №2 «Модели и алгоритмы машинного обучения» Тема 2.1 Алгоритмы восстановления пропусков Тема 2.2 Задача понижения размерности Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии Тема 2.4 Логические методы классификации. Деревья решений Тема 2.5 Ансамбли моделей Тема 2.6 Задача кластеризации Тема 2.7 Обучение с под-</p>	<p>ElasticNet. Сопоставление данной модели с моделями Ridge и LASSO. Логистическая регрессия. Сигмоидная функция в задачах бинарной классификации. Метод опорных векторов (SVM). Понятие отступа и постановка оптимизационной задачи. Аппроксимация и регуляризация эмпирического риска. Метод опорных векторов (SVM). Обобщение метода на нелинейные классификаторы. Логические методы классификации. Деревья решений. Алгоритм построения решающих деревьев ID3. Обработка пропусков в решающих деревьях. Переусложнение структуры решающего дерева. Прунинг. Деревья для регрессии. Алгоритм CART.</p>	<p>Анализ проявленных умений при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, анализ докладов на семинарских занятиях; анализ защиты рефератов; ана-</p>

	креплением		лиз решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ
ИПК 3.3. Владеет основами выявления первоначальных требований заказчика к ИС; системой информирования заказчика о возможностях типовой ИС и вариантах ее модификации; определениями возможностей достижения соответствия ИС первоначальным требованиям заказчика; основами составления протоколов переговоров с заказчиком	<p>Раздел 1. «Введение в машинное обучение»</p> <p>Тема 1.1 Основные понятия и постановка задачи машинного обучения. Этапы решения задач методами машинного обучения</p> <p>Тема 1.2 Предварительная обработка данных. Проектирование признаков</p> <p>Раздел №2 «Модели и алгоритмы машинного обучения»</p> <p>Тема 2.1 Алгоритмы восстановления пропусков</p> <p>Тема 2.2 Задача понижения размерности</p> <p>Тема 2.3 Наивный байесовский классификатор. Метрические методы классификации и регрессии</p> <p>Тема 2.4 Логические методы классификации. Деревья решений</p> <p>Тема 2.5 Ансамбли моделей</p> <p>Тема 2.6 Задача кластеризации</p> <p>Тема 2.7 Обучение с подкреплением</p>	<p>Композиции алгоритмов. Простое голосование базовых классификаторов. Методы повышения различности классификаторов.</p> <p>Случайный лес.</p> <p>Градиентный бустинг. Алгоритм AdaBoost.</p> <p>Искусственные нейронные сети в решении задач машинного обучения. Линейная модель нейрона.</p> <p>Нейронная реализация логических функций.</p> <p>Градиентный спуск для обучения искусственных нейронных сетей. Алгоритм обратного распространения ошибки.</p> <p>Понятие глубоких нейронных сетей. Сверточные, рекуррентные, рекурсивные нейронные сети.</p> <p>Постановка задачи кластеризации. Типы кластерных структур. Метод k-means.</p> <p>Агломеративная иерархическая кластеризация. Визуализация кластерной структуры.</p> <p>Постановка задачи частичного обучения. Применение классификации и кластеризации для решения задач частичного обучения.</p> <p>Методы понижения размерности. Метод главных компонент</p> <p>Методы понижения размерности. Нелинейные методы понижения размерности.</p> <p>Обучения с подкреплением: понятие и прикладное значение.</p> <p>Классификация методов машинного обучения применительно к типам решаемых задач.</p>	<p>Анализ проявленных навыков при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ</p>

## 7.2. Перечень вопросов для подготовки к промежуточной аттестации (зачет)

1. Понятие машинного обучения. Типы машинного обучения и виды задач, решаемых в рамках данного направления.
2. Этапы решения задач методами машинного обучения. Межотраслевой стандарт решения задач интеллектуального анализа данных.
3. Постановка задачи машинного обучения с учителем. Типы признаков.
4. Понятие эмпирического риска и сведение задачи обучения к задаче оптимизации.
5. Примеры задач машинного обучения в различных прикладных областях (медицина, финансовая сфера и т.д.)



6. Особенности сбора данных для решения задач машинного обучения: число признаков, количество и качество обучающей выборки.
7. Предварительная обработка данных. Обработка категориальных признаков.
8. Предварительная обработка данных. Обработка отсутствующих значений.
9. Предварительная обработка данных. Нормализация признаков.
10. Предварительная обработка данных. Стратегии выбора признаков.
11. Визуализация взаимосвязей между признаками и целевой переменной.
12. Основы проектирования признаков.
13. Усовершенствованное проектирование признаков. Признаки, извлекаемые из текстов и изображений.
14. Проблема переобучения.
15. Методы оценки модели. Оценка на отложенных данных.
16. Методы оценки модели. Кросс-валидация по k-блокам.
17. Оценка моделей бинарной классификации. Таблицы сопряженности. ROC-кривая.
18. Оценка моделей многоклассовой классификации. Таблицы сопряженности. ROC-кривые.
19. Оценка моделей регрессии.
20. Оптимизация модели путем подбора параметров. Алгоритм сеточного поиска.
21. Байесовская классификация. Гауссов и полиномиальный наивный байесовский классификаторы.
22. Метрические методы классификации и регрессии. Метод k-ближайших соседей (kNN).
23. Метрические методы классификации и регрессии. Метод k-ближайших соседей (kNN). Метод окна Парзена.
24. Метрические методы классификации и регрессии. Метрические классификаторы в задачах восстановления регрессии. Переход к непараметрической регрессии.
25. Формальная постановка задачи линейной регрессии и классификации. Наиболее часто используемые функции для аппроксимации пороговой функции.
26. Градиентный метод численной минимизации в задачах линейной классификации. Метод стохастического градиента.
27. Линейные модели регрессии. Применение линейной регрессии к нелинейным отношениям между переменными.
28. Проблема мультиколлинеарности. Разновидности линейных моделей и применение регуляризаторов.
29. Гребневая регрессия (Ridge, L2-регуляризация). LASSO (L1-регуляризация).
30. ElasticNet. Сопоставление данной модели с моделями Ridge и LASSO.
31. Логистическая регрессия. Сигмоидная функция в задачах бинарной классификации.
32. Метод опорных векторов (SVM). Понятие отступа и постановка оптимизационной задачи. Аппроксимация и регуляризация эмпирического риска.
33. Метод опорных векторов (SVM). Обобщение метода на нелинейные классификаторы.
34. Логические методы классификации. Деревья решений. Алгоритм построения решающих деревьев ID3.
35. Обработка пропусков в решающих деревьях.
36. Переусложнение структуры решающего дерева. Прунинг.
37. Деревья для регрессии. Алгоритм CART.
38. Композиции алгоритмов. Простое голосование базовых классификаторов. Методы повышения различности классификаторов.
39. Случайный лес.
40. Градиентный бустинг. Алгоритм AdaBoost.
41. Искусственные нейронные сети в решении задач машинного обучения. Линейная модель нейрона. Нейронная реализация логических функций.
42. Градиентный спуск для обучения искусственных нейронных сетей. Алгоритм обратного распространения ошибки.

43. Понятие глубоких нейронных сетей. Сверточные, рекуррентные, рекурсивные нейронные сети.
44. Постановка задачи кластеризации. Типы кластерных структур. Метод k-means.
45. Агломеративная иерархическая кластеризация. Визуализация кластерной структуры.
46. Постановка задачи частичного обучения. Применение классификации и кластеризации для решения задач частичного обучения.
47. Методы понижения размерности. Метод главных компонент
48. Методы понижения размерности. Нелинейные методы понижения размерности.
49. Обучения с подкреплением: понятие и прикладное значение.
50. Классификация методов машинного обучения применительно к типам решаемых задач.

### **7.3. Описание показателей и критериев оценивания сформированности компетенций на различных этапах их формирования; шкалы и процедуры оценивания**

#### **7.3.1. Оценивание ответов на вопросы и выполнения заданий для текущей и промежуточной аттестации**

При оценке знаний учитывается уровень сформированности компетенций:

1. Уровень усвоения теоретических положений дисциплины, правильность формулировки основных понятий и закономерностей.
2. Уровень знания фактического материала в объеме программы.
3. Логика, структура и грамотность изложения вопроса.
4. Умение связать теорию с практикой.
5. Умение делать обобщения, выводы.

#### **Шкала оценивания на зачете**

<b>Оценка</b>	<b>Критерии выставления оценки</b>
зачтено	Обучающийся должен: - продемонстрировать общее знание изучаемого материала; - показать общее владение понятийным аппаратом дисциплины; - уметь строить ответ в соответствии со структурой излагаемого вопроса; - знать основную рекомендуемую программой учебную литературу.
не зачтено	Обучающийся демонстрирует: - незнание значительной части программного материала; - не владение понятийным аппаратом дисциплины; - существенные ошибки при изложении учебного материала; - неумение строить ответ в соответствии со структурой излагаемого вопроса; - неумение делать выводы по излагаемому материалу

### **7.4. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРЫ ОЦЕНИВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ.**

Качество знаний характеризуется способностью обучающегося точно, структурированно и уместно воспроизводить информацию, полученную в процессе освоения дисциплины, в том виде, в котором она была изложена в учебном издании или преподавателем.

Умения, как правило, формируются на занятиях семинарского типа. Задания, направленные на оценку умений, в значительной степени требуют от обучающегося проявления стереотипности мышления, т.е. способности выполнить работу по образцам, с которыми он работал в процессе обучения. Преподаватель же оценивает своевременность и правильность выполнения задания.

Навыки можно трактовать как автоматизированные умения, развитые и закрепленные осознанным самостоятельным трудом. Навыки формируются при самостоятельном выполнении обучающимися практико-ориентированных заданий, моделирующих решение им производственных и социокультурных задач в соответствующей области профессиональной деятельности, как правило, при выполнении до-

машинных заданий, курсовых проектов (работ), научно-исследовательских работ, прохождении практик, при работе индивидуально или в составе группы и т.д.

Устный опрос – это процедура, организованная как специальная беседа преподавателя с группой обучающихся (фронтальный опрос) или с отдельными обучающимися (индивидуальный опрос) с целью оценки сформированности у них основных понятий и усвоения учебного материала. Устный опрос может использоваться как вид контроля и метод оценивания формируемых компетенций (как и качества их формирования) в рамках самых разных форм контроля, таких как: собеседование, коллоквиум, зачет, экзамен по дисциплине. Устный опрос (УО) позволяет оценить знания и кругозор обучающегося, умение логически построить ответ, владение монологической речью и иные коммуникативные навыки. УО обладает большими возможностями воспитательного воздействия преподавателя. Воспитательная функция УО имеет ряд важных аспектов: профессионально-этический и нравственный аспекты, дидактический (систематизация материала при ответе, лучшее запоминание материала при интеллектуальной концентрации), эмоциональный (радость от успешного прохождения собеседования) и др. Обучающая функция УО состоит в выявлении деталей, которые по каким-то причинам оказались недостаточно осмысленными в ходе учебных занятий и при подготовке к зачёту или экзамену. УО обладает также мотивирующей функцией: правильно организованные собеседование, коллоквиум, зачёт и экзамен могут стимулировать учебную деятельность студента, его участие в научной работе.

Тесты являются простейшей формой контроля, направленной на проверку владения терминологическим аппаратом, современными информационными технологиями и конкретными знаниями в области фундаментальных и прикладных дисциплин. Тест может предоставлять возможность выбора из перечня ответов (один или несколько правильных ответов).

Семинарские занятия. Основное назначение семинарских занятий по дисциплине – обеспечить глубокое усвоение обучающимися материалов лекций, прививать навыки самостоятельной работы с литературой, воспитывать умение находить оптимальные решения в условиях изменяющихся отношений, формировать современное профессиональное мышление обучающихся. На семинарских занятиях преподаватель проверяет выполнение самостоятельных заданий и качество усвоения знаний, умений, определяет уровень сформированности компетенций.

Коллоквиум может служить формой не только проверки, но и повышения производительности труда студентов. На коллоквиумах обсуждаются отдельные части, разделы, темы, вопросы изучаемого курса, обычно не включаемые в тематику семинарских и других практических учебных занятий, а также рефераты, проекты и иные работы обучающихся.

Доклад, сообщение – продукт самостоятельной работы студента, представляющий собой публичное выступление по представлению полученных результатов решения определенной учебно-практической, учебно-исследовательской или научной темы.

Контрольная работа - средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу.

Профессионально-ориентированное эссе – это средство, позволяющее оценить умение обучающегося письменно излагать суть поставленной проблемы, самостоятельно проводить анализ этой проблемы с использованием аналитического инструментария соответствующей дисциплины, делать выводы, обобщающие авторскую позицию по поставленной профессионально-ориентированной проблеме.

Реферат – продукт самостоятельной работы студента, представляющий собой краткое изложение в письменном виде полученных результатов теоретического анализа определенной научной (учебно-исследовательской) темы, где автор раскрывает суть исследуемой проблемы, приводит различные точки зрения, а также собственные взгляды на нее.

Ситуационный анализ – это комплексный анализ ситуации, имевший место в реальной практике профессиональной деятельности специалистов. Комплексный анализ включает в себя следующие составляющие: причинно-следственный анализ (установление причин, которые привели к возникновению данной ситуации, и следствий ее развертывания), системный анализ (определение сущностных предметно-содержательных характеристик, структуры ситуации, ее функций и др.), ценностно-мотивационный анализ (построение системы оценок ситуации, ее составляющих, выявление мотивов, установок, позиций действующих лиц); прогностический анализ (разработка перспектив развития событий по позитивному и негативному сценарию), рекомендательный анализ (выработка рекомендаций относительно поведения действующих лиц ситуации), программно-целевой анализ (разработка программ деятельности для разрешения данной ситуации).

Творческое задание – это частично регламентированное задание, имеющее нестандартное решение и позволяющее диагностировать умения интегрировать знания различных научных областей, аргументировать собственную точку зрения, доказывать правильность своей позиции. Может выполняться в индивидуальном порядке или группой обучающихся.

Деловая и/или ролевая игра – совместная деятельность группы обучающихся и преподавателя под управлением преподавателя с целью решения учебных и профессионально-ориентированных задач путем игрового моделирования реальной проблемной ситуации. Позволяет оценивать умение анализировать и решать типичные профессиональные задачи.

«Круглый стол», дискуссия – интерактивные оценочные средства, позволяющие включить обучающихся в процесс обсуждения спорного вопроса, проблемы и оценить их умение аргументировать собственную точку зрения. Занятие может проводить по традиционной (контактной) технологии, либо с использованием телекоммуникационных технологий.

Проект – конечный профессионально-ориентированный продукт, получаемый в результате планирования и выполнения комплекса учебных и исследовательских заданий. Позволяет оценить умения обучающихся самостоятельно конструировать свои знания в процессе решения практических задач и проблем, ориентироваться в информационном пространстве и уровень сформированности аналитических, исследовательских навыков, навыков практического и творческого мышления. Может выполняться в индивидуальном порядке или группой обучающихся.

## **РАЗДЕЛ 8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

При организации процесса изучения дисциплины «Избранные вопросы анализа данных» обучающемуся рекомендуется придерживаться следующих указаний:

1. Ознакомиться с общим тематическим планом лекционных и лабораторных занятий;
2. Самостоятельно прорабатывать уже пройденный материал лекционных и лабораторных занятий. При необходимости составить список вопросов и обратиться к преподавателю;
3. Перед изучением нового теоретического материала желательно заранее ознакомиться с содержанием предстоящей лекции. При необходимости составить список вопросов и обратиться к преподавателю.
4. Если в ходе рассмотрения нового теоретического или практического материала преподаватель ссылается на полученные ранее знания, умения или навыки, то рекомендуется их повторить;
5. Выполнять индивидуальные задания желательно не только в рамках учебных занятий;
6. Ознакомиться со списком литературы и интернет-ресурсов, рекомендуемых преподавателем для углубленного изучения либо дисциплины в целом, либо отдельных разделов.

Учебно-методическое обеспечение самостоятельной работы студентов по дисциплине включает:

1. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/26444.html>
2. Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 308 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/26445.html>

## **РАЗДЕЛ 9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### *Основная литература<sup>1</sup>*

1. Аверченков, В. И. Эволюционное моделирование и его применение : монография / В. И. Аверченков, П. В. Казаков. — Брянск : Брянский государственный технический университет, 2012. — 200 с. — ISBN 5-89838-441-X. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/7012.html>
2. Информационные технологии : учебник / Ю. Ю. Громов, И. В. Дидрих, О. Г. Иванова [и др.]. — Тамбов : Тамбовский государственный технический университет, ЭБС АСВ, 2015. — 260 с. — ISBN 978-5-8265-1428-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/63852.html>

### *Дополнительная литература<sup>2</sup>*

3. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. —

---

<sup>1</sup> Из ЭБС

<sup>2</sup> Из ЭБС

ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/45418.html>

4. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. — Новосибирск : Новосибирский государственный технический университет, 2011. — 888 с. — ISBN 978-5-7782-1590-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprsmarthop.ru/47719.html>

*Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине: интернет-ресурсы, современные профессиональные базы данных, информационные справочные системы*

#### **Интернет-ресурсы**

URL: <https://www.iprsmarthop.ru/> – электронно-библиотечная система Iprsmart.

#### **Информационно-справочные и поисковые системы**

Справочная правовая система «КонсультантПлюс»: <http://www.con-sultant.ru>

#### **Современные профессиональные базы данных**

URL:<http://www.edu.ru/> – библиотека федерального портала «Российское образование»

URL:<http://www.prlib.ru> – Президентская библиотека

URL:<http://www.rusneb.ru> – Национальная электронная библиотека

URL:<http://elibrary.rsl.ru/> – сайт Российской государственной библиотеки (раздел «Электронная библиотека»)

URL:<http://elib.gnpbu.ru/> – сайт Научной педагогической электронной библиотеки им. К.Д. Ушинского

*Комплект лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства*

#### **Комплект лицензионного программного обеспечения**

Microsoft Open Value Subscription для решений Education Solutions № Tr000544893 от 21.10.2020 г. MDE Windows, Microsoft Office и Office Web Apps. (срок действия до 01.11.2023 г.)

Антивирусное программное обеспечение ESET NOD32 Antivirus Business Edition договор № ИС00-006348 от 14.10.2022 г. (срок действия до 13.10.2025 г.)

Программное обеспечение «Мираполис» система вебинаров - Лицензионный договор 244/09/16-к от 15.09.2016 (Спецификация к Лицензионному договору 244/09/16-к от 15.09.2016, от 11.05.2022 г. ) (срок действия до 10.07.2023 г.)

Электронная информационно-образовательная среда «1С: Университет» договор от 10.09.2018 г. №ПРКТ-18281 (бессрочно)

Информационная система «ПервыйБит» сублицензионный договор от 06.11.2015 г. №009/061115/003 (бессрочно)

Система тестирования Indigo лицензионное соглашение (Договор) от 08.11.2018 г. №Д-54792 (бессрочно)

Информационно-поисковая система «Консультант Плюс» - договор об информационно поддержке от 26.12.2014, (бессрочно)

Электронно-библиотечная система IPRsmart лицензионный договор от 01.09.2021 г. №8234/21С (срок действия до 31.08.2024 г.)

Научная электронная библиотека eLIBRARY лицензионный договор SCIENC INDEX № SIO - 3079/2022 от 12.01.2022 г. (срок действия до 27.01.2024 г.)

#### **Свободно распространяемое программное обеспечение**

Комплект онлайн сервисов GNU ImageManipulationProgram, свободно распространяемо программное обеспечение

Веб-браузер, Google Chrome, свободное ПО, ежегодно обновляемое ПО.

Пакет офисных приложений, Office 2016, лицензионное соглашение - Договор №Tr000544893 от 21/10/2020 – 3 года

Пакет офисных приложений, OpenOffice, свободное ПО, ежегодно обновляемое ПО

Просмотр файлов в формате PDF, Adobe Reader, свободно распространяемое ПО, ежегодно обновляемое ПО

Просмотр файлов в формате DJV, WinDjView, свободное ПО, ежегодно обновляемое ПО

Файловый архиватор, 7 Zip, свободное ПО, ежегодно обновляемое ПО

Файловый менеджер, Far, свободно распространяемое ПО, ежегодно обновляемое ПО

Anaconda: дистрибутив языков программирования Python и R.

### **Программное обеспечение отечественного производства:**

Программное обеспечение «Мираполис» система вебинаров - Лицензионный договор 244/09/16-к от 15.09.2016 (Спецификация к Лицензионному договору 244/09/16-к от 15.09.2016, от 11.05.2022 г. ) (срок действия до 10.07.2023 г.)

Электронная информационно-образовательная среда «1С: Университет» договор от 10.09.2018 г. №ПРКТ-18281 (бессрочно)

Информационная система «ПервыйБит» сублицензионный договор от 06.11.2015 г. №009/061115/003 (бессрочно)

Система тестирования Indigo лицензионное соглашение (Договор) от 08.11.2018 г. №Д-54792 (бессрочно)

Информационно-поисковая система «Консультант Плюс» - договор МИ-ВИП-79717-56/2022 от 23.12.2021 (срок действия до 31.12.2022 г.)

Информационно-поисковая система «Консультант Плюс» - договор об информационно поддержке от 26.12.2014, (бессрочно)

Электронно-библиотечная система IPRsmart лицензионный договор от 01.09.2021 г. №8234/21С (срок действия до 31.08.2024 г.)

Научная электронная библиотека eLIBRARY лицензионный договор SCIENC INDEX № SIO -3079/2022 от 12.01.2022 г. (срок действия до 27.01.2024 г.)

## **РАЗДЕЛ 10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	<b>Оборудование:</b> специализированная мебель (мебель аудиторная (9 столов, 9 стульев, доска аудиторная), стол преподавателя, стул преподавателя). <b>Технические средства обучения:</b> персональный компьютер преподавателя; 9 компьютеров, мультимедийное оборудование (проектор, экран).
Помещение для самостоятельной работы	Специализированная мебель (10 столов, 10 стульев), персональные компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета