

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 24.02.2026 15:41:33
Уникальный программный ключ:
880f7c07c583b07b775f6604a630216c5a911

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

СЕВЕРО-ЗАПАДНЫЙ ИНСТИТУТ УПРАВЛЕНИЯ

УТВЕРЖДЕНА
Решением УС СЗИУ РАНХиГС
от «17» февраля 2026 г. протокол № 5

**ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
повышения квалификации
ПРИКЛАДНАЯ СТАТИСТИКА И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ
ДАННЫХ**

Санкт-Петербург, 2026

Разработчик:

Доктор военных наук, профессор
Заведующий кафедрой бизнес-информатики
(ученая степень и (или) ученое звание, должность,
структурное подразделение)

В.Н. Наумов
(И.О. Фамилия)

Руководитель структурного подразделения

кандидат политических наук, доцент, декан ФДПО
(ученая степень и (или) ученое звание, должность, структурное подразделение)



Н.В. Горбатова
(И.О. Фамилия)

Программа повышения квалификации рассмотрена на заседании ученого совета СЗИУ и рекомендована к реализации, протокол № 1 от «27» января 2026 г.

СОДЕРЖАНИЕ

1 ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ.....	4
1.1. Цель и задачи реализации программы	4
1.2. Нормативная правовая база.....	4
1.3. Планируемые результаты обучения	5
1.4. Категория слушателей	7
1.5. Формы обучения и сроки освоения	7
1.6. Период обучения и режим занятий	7
1.7. Документ о квалификации	7
2 СОДЕРЖАНИЕ ПРОГРАММЫ	7
2.1. Календарный учебный график.....	7
2.2. Учебный план	8
2.3. Содержание программы по темам.....	9
3 ОРГАНИЗАЦИОННЫЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ	11
3.1. Материально-техническое и программное обеспечение реализации программы.....	11
3.2. Учебно-методическое и информационное обеспечение программы	12
4 ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ.....	21
5 ИНДИКАТОРЫ СФОРМИРОВАННЫХ КОМПЕТЕНЦИЙ ВЫПУСКНИКА ПРОГРАММЫ (при необходимости)	26

1 ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1. Цель и задачи реализации программы

Цель дополнительной профессиональной программы повышения квалификации «Прикладная статистика и интеллектуальный анализ данных» (далее – Программа) – совершенствование компетенций, направленных на решение задач прикладной статистики и интеллектуального анализа данных, использование технологий анализа больших данных, машинного обучения, искусственного интеллекта при решении практических задач анализа, моделирования и прогнозирования в предметной области в условиях больших данных.

В процессе обучения должны быть решены следующие задачи:

- ознакомление с базовыми методами подготовки и исследования данных;
- ознакомление с современными методами и инструментальными средствами инструментальными средствами анализа данных и машинного обучения;
- формирование начальных навыков в решении частных задач разведочного анализа данных, многомерной статистики, интеллектуального анализа данных с применением современных программных средств прикладной статистики, машинного обучения, интеллектуального анализа данных.

1.2. Нормативная правовая база

Программа разработана на факультете дополнительного профессионального образования. на основании ряда законов и нормативных правовых актов в области дополнительного профессионального образования, в т.ч:

Федеральный закон от 29.12.2012 N 273-ФЗ «Об образовании в Российской Федерации»;

Постановление Правительства РФ от 12.05.2012 N 473 (ред. от 30.07.2025) «Об утверждении устава федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации»;

Приказ Минобрнауки России от 01.07.2013 N 499 (ред. от 15.11.2013) «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам» (Зарегистрировано в Минюсте России 20.08.2013 N 29444);

Приказ Минобрнауки России от 23.08.2017 N 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;

Приказ РАНХиГС Об утверждении новой редакции образовательного стандарта академии по направлению подготовки 38.03.05 «Бизнес-информатика» от 01 сентября 2023 года № 01-23812;

Приказ Минтруда России от 06.07.2020 N 405н "Об утверждении профессионального стандарта "Специалист по большим данным" (Зарегистрировано в Минюсте России 05.08.2020 N 59174)

Приказ Минобрнауки России от 24 марта 2025 года № 266 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам» (Зарегистрировано в Минюсте России 22 апреля 2025 года, рег. № 81928);

Приказ РАНХиГС от 02 декабря 2025 года № 02-0 2669/001 «Об утверждении порядка разработки и утверждения в Академии дополнительных профессиональных программ – программ повышения квалификации, программ профессиональной переподготовки»

1.3. Планируемые результаты обучения

Планируемые результаты обучения включены в таблицу (таблица 1).

Таблица 1. Планируемые результаты обучения

Виды деятельности	Профессиональные компетенции или трудовые функции (формируются и (или) совершенствуются) УК	Знания	Умения	Практический опыт
ВД-1. Аналитическая	Профессиональный стандарт 06.042 – «специалист по большим данным» Обобщенная трудовая функция А.6 Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	Современные методы и инструментальные средства анализа больших данных; Теоретические и прикладные основы анализа больших данных	Обрабатывать эмпирические и экспериментальные данные Производить очистку данных для проведения аналитических работ	Извлечения, проверки и очистки данных построения и применения инструментальных средств обработки больших данных
	Трудовая функция А/03.6 Подготовка данных для проведения аналитических работ по исследованию больших данных	основные понятия и основные методы,	осуществлять предобработку и очистку данных, выполнять	использования информации, методов и программных

	<p>применением технологий больших данных в соответствии с требованиями заказчика</p>	<p>многомерной математической статистики Технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, планирование экспериментов, карты контроля качества</p>	<p>разведывательный анализ Проводить сравнительный анализ методов и инструментальных средств анализа больших данных Использовать имеющуюся у исполнителя методологическую и технологическую инфраструктуру анализа больших данных для выполнения аналитических работ</p>	<p>средств ее сбора, обработки информационно-аналитической поддержки принятия управленческих решений</p>
<p>Универсальные компетенции (формируются и (или) совершенствуются)</p>				
УК-1	<p>Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</p>	<p>Основных поисковых систем</p>	<p>Умеет применять системное и критическое мышление для решения поставленных задач</p>	<p>Владеет навыками анализа и синтеза информации</p>

1.4. Категория слушателей

Программа профессионального обучения разработана в рамках федерального проекта "Активные меры содействия занятости" национального проекта "Кадры".

Условиями участия отдельных категорий граждан в мероприятиях по обучению является отнесение их к одной из категорий, предусмотренных Постановлением Правительства РФ от 07.03.2025 N 291. "Об утверждении Положения о реализации мероприятий по организации профессионального обучения и дополнительного профессионального образования отдельных категорий граждан".

К освоению программы допускаются лица, имеющие и/или получающие высшее образование.

1.5. Формы обучения и сроки освоения

Форма обучения: очная

Срок освоения - 72 акад.ч.

Контактная работа – 48 акад.ч.

Самостоятельная работа – 22 акад.ч.

Итоговая аттестация – 2 акад.ч.

1.6. Период обучения и режим занятий

Продолжительность обучения – 9 дней.

Режим занятий – не более 8 акад. часов в день.

1.7. Документ о квалификации

Вид документа, выдаваемый при успешном освоении программы - удостоверение о повышении квалификации РАНХиГС.

2 СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. Календарный учебный график

Таблица 2. Календарный учебный график

	Период обучения - (9 дней)							
1	2	3	4	5	6	7	8	9
УЗ/СР	УЗ/СР	УЗ/СР	УЗ/СР	УЗ/СР	УЗ/СР	УЗ/СР	УЗ/СР	УЗ/ИА

УЗ – учебные занятия в очном формате

СР – самостоятельная работа

ИА - итоговая аттестация

2.2. Учебный план

Таблица 3. Учебный план

№п/п ¹	Наименование темы	Общая трудоемкость, час.	Контактная работа, час.					Самостоятельная работа, час ⁸	Контактная работа (с применением дистанционных образовательных технологий), час. ⁶					Самостоятельная работа (в т.ч. электронное обучение (ЭО), час ⁷	Текущий контроль успеваемости ⁸	Промежуточная аттестация (форма/час) ⁹	Итоговая аттестация (вид /час.) ¹⁰	Код компетенций ¹¹	
			Всего	В том числе					Всего ⁴	В том числе									
				Лекции / в интерактивной форме	Практические (семинарские/лабораторные) занятия /в интерактивной форме ⁶	Контактная самостоятельная работа, час ⁷	Индивидуальные и групповые консультации			Лекции/ в интерактивной форме ⁵	Практические (семинарские/лабораторные) занятия /в интерактивной форме ⁵	Контактная самостоятельная работа. час ⁷	Индивидуальные и групповые консультации						
1.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1.	Общая характеристика прикладной статистики и интеллектуального анализа данных	18	12	4	8			6											УК-1
2.	Подготовка данных для проведения аналитических работ	18	12	4	8			6											УК-1
3.	Многомерный статистический анализ данных	16	12	4	8			4											УК-1
4.	Интеллектуальный анализ данных и машинное обучение	18	12	4	8			6											УК-1
	Итого:	70	48	16	32	0	0	22	0	0	0	0	0	0	0	0	0	0	0
	Итоговая аттестация	2	2														2		
	Всего:	72	50	16	32	0	0	22	0	0	0	0	0	0	0	0	2	0	

1. 3 – зачет
 2. Т – традиционная форма

2.3. Содержание программы по темам

Приводится содержание программы. Содержание теоретического и практического материала раскрывается в логической последовательности изучения модулей/разделов/дисциплин/тем учебного плана. Содержание программы раскрывается с учетом современного развития образования и науки, техники, культуры, а также перспектив их развития (Таблица 4).

Таблица 4

**Содержание программы по модулям/разделам/дисциплинам/
и (или) темам**

Номер модуля/раздела/дисциплины/темы и его наименование	Содержание темы
Тема 1. Общая характеристика прикладной статистики и интеллектуального анализа данных	<p>Введение. Понятие анализа данных. Задачи систем поддержки принятия решений. Жизненный цикл процесса анализа данных. Классификация методов Data Mining. Модели Data Mining. Понятие данные и знания. Типы и шкалы данных. Репозитории данных. Классические наборы данных. Задачи анализа данных. Описательная аналитика, Диагностическая аналитика. Предсказательная и предписывающая аналитика. Процесс обнаружения знаний. Машинное обучение. Основные проблемы машинного обучения. Переобучение и недообучение модели. Проблема смещения и большой дисперсии. Классификация задач Data Mining. Методы анализа данных. Задачи классификации и регрессии. Средства анализа данных и машинного обучения. Использование статистических пакетов для интеллектуального анализа данных. Классификация источников больших данных. Характеристики больших данных. Сферы применения больших данных. Первичные источники больших данных. Вторичные источники больших данных. Открытые источники больших данных.</p> <p>Язык SQL. Системы управления базами данных. Озера и хранилища данных. NoSQL. Экосистемы больших данных. Инструменты ETL/BI при работе с большими данными.</p>
Тема 2. Подготовка данных для проведения аналитических работ	<p>Методология KDD. Задачи предобработки данных. Технология ETL. Разведочный анализ данных. Модель анализа данных Тьюки. Набор данных Энскомба. Просмотр данных. Очистка данных. Оценка качества данных. Заполнение пропущенных данных. Аномальные и предельные данные. Использование ящичной диаграммы. Выявление дубликатов и противоречий. Корреляционный анализ. Трансформация данных. Квантование данных (биннинг). Сэмплинг. Повторные выборки. Бутстреп. Кроссвалидация. Перестановочный тест. Метод «складного ножа». Группировка данных. Решение задач описательной статистики. Графический анализ данных. Статистические диаграммы.</p> <p>Основные положения непараметрической и нечисловой статистики. Таблицы сопряженности. Таблица сопряженности 2x2. Таблицы флагов и заголовков. Непараметрические и нечисловые критерии. Корреляционная матрица. Задачи ассоциации.</p>

Номер модуля/раздела/ дисциплины/те- мы и его наименование	Содержание темы
<p>Тема 3. Многомерный статистический анализ данных</p>	<p>Агрегирование и группирование данных. Понятие проекции. Принцип ординации наблюдений. Задачи снижения размерности данных. Метод главных компонент. Организация решения задачи методом главных компонент. Матрица факторных нагрузок. Собственные числа. Критерий КМО. Критерий Кайзера, Критерий каменистой осыпи.</p> <p>Объясненная совокупная дисперсия. Факторный анализ. Основные этапы факторного анализа. Общность и характерность переменных. График факторных нагрузок. Методы вращения. анализа в различных программных средах. Оценка качества регрессионной модели. Проблемы мультиколлинеарности. Аномалии, влиятельные и напряженные наблюдения.</p> <p>Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Дендрограммы. Метод К-средних. Понятие центроида. Профили кластеров. Взаимосвязь кластерного и регрессионного анализа. Задачи регрессионного анализа. Парные и множественные регрессионные модели. Оценка качества моделей</p>
<p>Тема 4. Интеллектуаль- ный анализ данных и машинное обучение</p>	<p>Понятие Data Mining и машинного обучения. Процесс KDD Формулировка задачи классификации. Классификационный анализ с обучением. Метод k-ближайших соседей. Наивный байесовский классификатор. Логистическая регрессия. Деревья решений. Метод опорных векторов. Использование нейронных сетей для решения задач классификации. Ансамблевые методы классификации. Сравнение результатов классификации различными методами. Верификация. Оценка качества классификации. ROC-кривая. Показатель AUC. Таблица сопряженности (матрица путаницы). Показатели точности: accuracy, recall, precision, F1. Проблема сбалансированности данных. Примеры решения задач интеллектуального анализа данных и машинного обучения. Использование статистических пакетов Logitom, JASP, Excel (R, Python).</p>

3 ОРГАНИЗАЦИОННЫЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

3.1. Материально-техническое и программное обеспечение реализации программы

Программа обеспечена оборудованными аудиториями, оснащёнными мультимедийным/видеопроекционным оборудованием, позволяющим работать с текстом, изображениями, воспроизводить демонстрационные материалы, в ходе проведения лекционных и практических занятий, текущего контроля успеваемости и итоговой аттестации.

Программа обеспечена условиями для функционирования электронной информационно-образовательной среды, включающей в себя лицензионные программные продукты Microsoft Office (Excel, Word, Outlook, Power Point и др), обеспечивающие освоение слушателями образовательной программы в полном объеме.

В образовательной деятельности предусмотрены следующие виды учебных занятий и учебных работ: лекции, практические занятия, включающие в т.ч. разбор кейсов, консультации, обеспечивающие высокое качество учебного процесса.

Темы занятий, даты и время проведения, а также преподаватели, задействованные в их проведении, указываются в программе (брошюра).

Обязательным условием проведения занятий выступает выделение 70% учебного времени на проведение практических занятий с использованием интерактивных образовательных технологий (практикумы и др.). Предусмотрена организация консультационной помощи слушателям.

На лекционных занятиях рассматриваются ключевые и наиболее сложные вопросы дисциплины. Лекция сопровождается презентациями, что позволяет самостоятельно работать над повторением и закреплением материала.

Практические занятия проводятся в компьютерных классах с рабочими местами, имеющими выход в сеть Интернет. Тематика практических работ направлена на закрепление и углубление теоретических знаний, полученных на лекциях, на подготовку к дальнейшей работе по дисциплине, на выработку навыков и получение опыта для реальной работы.

Подготовка к практической работе предусматривает изучение теоретического материала. Перед выполнением практической работы необходимо внимательно ознакомиться с описанием практического задания, уяснить, в чем состоят её цель и заданные результаты. Выполнение каждой работы сопровождается оформлением в виде слайдов презентации.

Для активизации работы слушателей во время контактной работы с преподавателем часть занятий проводятся в интерактивной форме.

3.2. Учебно-методическое и информационное обеспечение программы

Тема 1. Общая характеристика прикладной статистики и интеллектуального анализа данных

Входной тест для проверки готовности слушателей к занятиям и адаптации содержания и стиля предъявления материалов

1.Какой характеристикой оценивается мера рассеивания случайной величины?

- A) Модой
- B) Асимметрией
- C) Дисперсией

ANSWER: C

2.Как увеличить точность оценки неизвестной характеристики?

- A) Использовать другой метод измерения значения характеристики
- B) Увеличить число наблюдений
- C) Выбрать другую характеристику

ANSWER: B

3.Что оценивает гистограмма распределения?

- A) Текущее значение наблюдения
- B) Плотность распределения
- C) Функцию распределения

ANSWER: B

4.Что такое выборка?

- A) Набор predetermined данных
- B) Множество вариантов решения задачи
- C) Конечное число результатов наблюдения, по которой анализируется

вся совокупность

ANSWER: C

Тест по теме 1

1. Какой оператор используется для проверки существования записи?

- a) EXISTS;
- b) IN;
- c) ANY;
- d) HAS;

ANSWER: A

2. Интерпретировать понятие реляционной модели «Атрибут»

- a) Строка таблицы
- b) Заголовок столбца таблицы
- c) Заголовок таблицы
- d) Столбец таблицы

ANSWER: D

3. Определить для чего используется схема данных

- a) Создание связи
- b) Поиск данных
- c) Отображение данных
- d) Фильтрация данных

ANSWER: A

Тема 2. Подготовка данных для проведения аналитических работ

Задания по теме 2

Задание 1. Решить задачу разведочного анализа для набора данных Boston.

Задание 2. Набор данных имеет пропуски и аномальные значения

X	Y
1	1
-1	-3
3	4
2	2
5	7
2	2
3	4
5	7
1	1
3	4
4	25
4	5
5	7
1	1
3	4
3	4
2	2
4	6
0	-1
3	4
4	5
7	10
1	1
7	10
3	4
3	4
3	3
4	5
4	19
4	6

1	0
-4	-23
4	6
1	1
2	2
5	7
1	39
4	5
5	7
7	11
4	6
6	8
4	5
2	3
4	6
1	1
5	7
5	7
5	7
3	4
-1	-3
4	6
8	12
3	4
6	9
6	9
4	5
5	7
8	12
3	4
4	6
3	4
3	4
6	9
1	1
4	
4	6
4	5
5	7
5	7
1	1
2	2
2	2
6	9

4	5
2	2
5	
1	1
2	2
1	1
5	7
5	7
2	2
-1	
3	3
3	4
4	5
7	10
6	8
5	7
3	4
4	6
4	5
6	9
5	7
5	
4	5
4	5
1	0
1	1

1. Проверить гипотезу о нормальном законе распределения для переменных X,Y
2. Решить задачу восстановления пропущенных данных;
3. Найти аномальные значения.
4. Проанализировать корреляцию между переменными. Проверить гипотезу о значимости коэффициента корреляции

Тема 3. Многомерный статистический анализ данных

Задание 1.

Имеется выборка данных о 6 предприятиях. Найти расстояние между объектами с помощью различных метрик

	X1	X2	X3
1	120	9100	11000
2	180	8400	16000
3	840	13000	20000
4	410	11300	16000
5	460	12000	15000

6	560	11500	13000
---	-----	-------	-------

С помощью принципа "ближайшего соседа" для метрики Евклида построить дендрограмму. Задачу решить в различных приложениях. Построить workflow-diagram Orange

Задание 2.

На предприятии существуют 5 отделов. Поскольку в них имеется разное число сотрудников, разные виды деятельности и др. решено сгруппировать отделы. Решить задачу группирования иерархическим и методом k-средних.

	Стоимость производственных фондов, X1	Среднемесячный объем работ, X2
1	699	190
2	510	210
3	340	110
4	290	95
5	310	130

Задачу решить в JASP

Тесты по теме 3

Результаты иерархического кластерного анализа, как правило, представляются в виде:

- A) дендрограммы
- B) дерева решений
- C) блок-схемы алгоритма

ANSWER: A

При решении задачи кластерного анализа для набора данных iris различными методами построены дендрограммы с выделением кластеров прямоугольниками для заданного числа кластеров равного трем. Выберите лучший из методов, если дендрограммы имеют следующий вид

- A) ближнего соседа
- B) дальнего соседа
- C) центроидный
- D) средней связи
- E) взвешенной средней связи

ANSWER: B

Как называется метод, в начале выполнения которого каждое наблюдение представляет собой отдельный кластер. На каждом шаге его выполнения два ближайших кластера объединяются в новый кластер

- A) агломеративный
- B) дивизимный
- C) k-средних
- D) метод меноидов

ANSWER: A

Какой из методов кластерного анализа при анализе кандидатов на включения в кластер на текущем шаге использует результаты оценки дисперсий?

- A) метод Варда
- B) метод ближайшего соседа
- C) метод полной связи
- D) центроидный метод
- E) метод невзвешенного попарного среднего

ANSWER: A

Что входит в состав дерева решений?

- A) Отдельные наблюдения
- B) Примеры вариантов решений
- C) Множество вершин со значениями атрибутов и с вариантами решений

ANSWER: C

Тема 4. Интеллектуальный анализ данных и машинное обучение

Задание 1. Построить дерево решений по данным, приведенным в таблице.

Рейтинг	Возраст	Уровень Дохода	Образование
0	35	3000	0
0	25	5000	1
0	31	7000	1
1	56	1000	0
1	62	1100	1
1	49	1500	0

Задание 2. Решить задачу логистической регрессии. Определить качество построенной модели классификации. Решить данную задачу другим методами классификации, реализованными в Deductor Academic. Сравнить результаты решения задачи классификации с помощью таблицы сопряженности.

Рейтинг	Образование, A1	Доход, A2	Возраст, A3
низкий	высшее	малый	35
низкий	среднее	большой	40
высокий	высшее	большой	30
высокий	высшее	большой	30
низкий	среднее	малый	30
высокий	высшее	малый	35
высокий	высшее	большой	45
высокий	высшее	большой	35

Задание 3

Решить задачу классификации для предложенного набора данных. Имеется 15 наборов данных, размещенных в Moodle. Каждый студент решает

задачу классификации для своего варианта исходных данных. При решении задачи классификации следует решать задачи классификации методами логистической регрессии, деревьев решений, k-ближайших соседей, случайного леса. В Moodle находится пример решения задачи и оформления отчета.

Тест по теме 4

1.Выбор классификатора. Использование различных методов классификации позволяет получить следующие таблицы сопряженности.

Метод линейного дискриминантного анализа

69535	442
4546	477

Метод k ближайших соседей при k=9

69938	39
4973	50

Логистическая регрессия

69770	207
5005	18

Метод опорных векторов

69882	103
4890	125

Выбрать лучший метод по показателю precision

- метод опорных векторов
- метод k-ближайших соседей
- логистическая регрессия
- линейный дискриминантный анализ

ANSWER: B

2.Использование различных методов классификации позволяет получить следующие таблицы сопряженности.

Метод линейного дискриминантного анализа

69535	442
.	477

Метод k ближайших соседей при k=9

69938	39
4973	50

Логистическая регрессия

69770	207
5005	18

Метод опорных векторов

69882	103
4890	125

Выбрать лучший метод по показателю ассурасу

- метод опорных векторов
- метод k-ближайших соседей
- логистическая регрессия
- линейный дискриминантный анализ
- Деревья решений.

ANSWER: D

Основная литература

1. Афанасьев, Владимир Николаевич. Анализ временных рядов и прогнозирование. - Саратов: Ай Пи Ар Медиа. – 310 с. Текст : электронный. - URL: <http://www.iprbookshop.ru/90196.html> (дата обращения: 12.11.2020). - Режим доступа: для авторизир. пользователей
2. Голоктионова Ю.Г., Ильминская С.А., Илюхина И.Б., Луговской А.М., Лисичкина Н.В. и др. Прогнозирование и планирование в экономике. - Москва: Прометей – 544 с. Текст : электронный. - URL: <http://www.iprbookshop.ru/94511.html> (дата обращения: 01.10.2020). - Режим доступа: для авторизир. пользователей
3. Мастицкий С. Э. (2020) Анализ временных рядов с помощью R. — Электронная книга, адрес доступа: <https://ranalytics.github.io/tsa-with-r>
4. Миркин, Борис Григорьевич. Базовые методы анализа данных – М.: Юрайт, 2024 – 297 с. Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262> (дата обращения: 27.04.2025)
5. Мхитарян В. С., Архипова М. Ю., Дуброва Т. А., Миронкина Ю. Н., Сиротин В. П. Анализ данных. – М.: Юрайт, 2023 – 490 с. Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166> (дата обращения: 27.04.2025)
6. О`Нил, Кэти. Data Science : Инсайдерская информация для новичков. Включая язык R : [пер. с англ.] – СПб. Питер. – 368 с. Текст: электронный. - URL: <http://new.ibooks.ru/bookshelf/359209/reading> (дата обращения: 25.01.2021)
7. Хайндман Р. Дж, Атанасопулос Дж. Прогнозирование: принципы и практика. [Электронный ресурс] –URL: <https://otexts.com/fpp3/>

Дополнительная литература

1. Гусарова Н.Ф, Анализ социальных сетей. Основные понятия и метрики. – СПб.: Университет ИТМО, 2016 – 67 с.
2. Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow : Концепции, инструменты и техники для создания интеллектуальных систем : полноцветное издание : перевод с английского - ПрМ.: Диалектика. -684 с.
3. Ланц Б. – Машинное обучение на R/ - СПб. : Питер. – 2020 – 464 с.
4. Principles of Econometrics with R [Электронный ресурс] – URL: <https://bookdown.org/ccolonescu/RPoE4/>
5. Люк Д. Анализ сетей (графов) в среде R. Руководство пользователя- М.: ДМК Пресс, 2017 – 250 с.
6. Маккинни У. Python и анализ данных. – М.: ДМК, 2015 – 482 с.
7. Нильсен Эйлин. Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение. –М.: ООО Диалектика – 2021 – 544 с.
8. Наумов Владимир Николаевич. Средства бизнес- аналитики: учеб. пособие / В. Н. Наумов ; Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2016. - 107 с.
9. Наумов В.Н. Анализ данных и машинное обучение: методы и инструментальные средства. Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2020. - 260 с.

Нормативные правовые документы и иная правовая информация

Не используются

Интернет-ресурсы

Обучающимся обеспечен доступ к материалам курса в СДО Академии <http://lms.ranepa.ru>, а также через сайт научной библиотеки к следующим подписным электронным ресурсам:

Русскоязычные ресурсы

- Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»
- Электронные учебники электронно-библиотечной системы (ЭБС) «Юрайт»
- Электронные учебники электронно-библиотечной системы (ЭБС) «Лань»
- Электронные учебники электронно-библиотечной системы (ЭБС) «ZNANIUM.COM»
- Электронные учебники электронно-библиотечной системы (ЭБС) «BOOK.RU»
- Электронные учебники электронно-библиотечной системы (ЭБС) «IPRSMART»

4 ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

Контроль знаний осуществляется по результатам освоения программы повышения квалификации.

Оценочные средства текущего контроля представлены выполнением контрольных заданий и текущих тестов. итоговая аттестация – зачет в форме выполнения итогового задания и выполнения теста. Контрольные задания и текущие тесты относятся к категории текущего контроля усвоения материала курса. Максимальная оценка за текущий контроль составляет 60 баллов.

Практическая работа выполняется после каждого учебного занятия по теме занятия в аудитории и во время часов. В зависимости от полноты и корректности выполненного задания начисляются баллы от 5 до 10. Выполненные работы по каждой теме проверяются преподавателем до начала следующей темы.

Текущий контроль успеваемости включает контроль усвоения материала на практических занятиях, выполнение домашних заданий.

Итоговая аттестация проходит на последнем занятии в форме зачета в компьютерном классе.

Максимальная оценка итоговой аттестации 40 баллов (20 баллов итоговый тест и 20 баллов итоговое контрольное задание).

Оценочные средства (итогового контроля)	Показатель оценивания (Что делает обучающийся (какие действия способен выполнить))	Критерий оценивания (Как (с каким качеством) выполняется действие).
Решение контрольного задания	Знание основных теоретических положений анализа данных и машинного обучения, основных возможностей инструментальных средств и методики их использования	В зависимости от полноты и корректности выполненного задания начисляются баллы не более 20
Выполнение теста)	Умение использовать современные информационные технологии, инструментальные средства при решении практических задач анализа данных и машинного обучения	Тест выполняется с использованием системы электронного обучения Moodle. Максимальная сумма баллов составляет 20 баллов

Типовые проверочные задания для самоподготовки обучающихся к аттестации

Номер вопроса	Вопрос	Возможные ответы	Правильный ответ
1.	Что такое OLAP-системы?	Системы работы с пользователями	
		Системы оценки параметров	

		Системы аналитической обработки	+
2.	Что такое Data Mining?	Процесс определения данных	
		Процесс извлечения знаний из данных	+
		Процесс очистки «сырых данных»	
		Процесс исключения противоречий	
3.	Указать, что не относится к методам Data Mining	Кластерный анализ	
		Деревья решений	
		Нейронные сети	
		Контент-анализ	+
4.	Что такое выборка?	Набор predetermined данных	
		Множество вариантов решения	
		Конечное число результатов наблюдения, по которой анализируется вся совокупность	+
5.	Что такое репрезентативная выборка?	Выборка бесконечного размера	
		Выборка, содержащая разные значения признака	
		Выборка, по которой можно судить о генеральной совокупности	+
6.	Какие шкалы не используются при оценке признака?	Номинальная	
		Сильная	+
		Порядковая	
		Интервальная	
7.	Какой характеристикой оценивается центр распределения?	Дисперсией	
		Средним арифметическим	+
		Стандартной ошибкой	
		Ассиметрией	
8.	Какой характеристикой оценивается мера рассеивания случайно величины?	Модой	
		Асимметрией	
		Дисперсией	+
9.		Использовать другой метод измерения	

	Как увеличить точность оценки неизвестной характеристики?	значения характеристики	
		Увеличить число наблюдений	+
		Выбрать другую характеристику	
10.	Что оценивает гистограмма распределения?	Текущее значение наблюдения	
		Плотность распределения	+
		Функцию распределения	
11.	Что позволяет оценить коэффициент парной корреляции?	Силу линейной связи	+
		Качественные различия признаков	
		Меру изменчивости признака	
12.	Каким не может быть значение коэффициента корреляции	нулевым	
		отрицательным	
		Больше единицы	+
13.	Как проверяется статистическая гипотеза?	Расчетом наблюдаемого значения критерия и сравнением с пороговым значением	+
		Перебором различных вариантов	
		Проведением дополнительного эксперимента	
14.	Что показывает уровень значимости гипотезы?	Ее важность для практики	
		Вероятность ошибки при принятии решения о справедливости гипотезы	+
		Важность основной гипотезы по сравнению с альтернативной	
15.	Для чего используется регрессионный анализ?	Для проверки наличия взаимосвязи между признаками	
		Для построения выборочной модели	+
		Для группирования элементов	
16.	Что может быть использовано для проверки качества регрессионной модели?	Коэффициент детерминации	+
		Ковариационный момент	

		Математическое ожидание отклика	
17.	Для чего используется дисперсионный анализ?	Для построения модели	
		Для исследования значимости влияния изменения факторов на зависимую переменную	+
		Для экспресс-анализа переменных	
18.	Как производится интервальная оценка неизвестного параметра?	Определением границ доверительного интервала	+
		Построением модели зависимости	
		Проведением множества измерений и оценкой закона распределения	
19.	Что означает эффективность оценки?	Наличие наименьшей ошибки оценки	+
		Минимальные затраты оценки	
		Достоверность выполненной оценки	
20.	Для чего предназначен кластерный анализ?	Для группирования объектов по каким-либо признакам	+
		Для классификации объекта	
		Для выполнения разведывательного анализа имеемой зависимости	
21.	Что представляет собой нейронная сеть?	Цепь отдельных элементов-нейронов	
		Сеть искусственных нейронов	+
		Сеть передачи данных	
22.	Что входит в состав дерева решений?	Отдельные наблюдения	
		Примеры вариантов решений	
		Множество вершин со значениями атрибутов и с вариантами решений	+

Итоговое контрольное задание

Задача 1. Решить задачу регрессионного анализа для набора данных `рост_вес.csv`. Данный набор состоит из 25000 наблюдений и размещен в дистанционной системе обучения moodle.

Задача 2. Решить задачу регрессионного анализа для набора данных `стоимость жилья.csv`. Данный набор состоит из 25000 наблюдений и размещен в дистанционной системе обучения moodle.

Задача 3. Решить задачу для набора данных `РегрессияЗанятие.csv`

x1	x2	x3	x4	y
7,68	10,03	3,36	-11,47	35,27
7,93	11,53	-1,8	-7,06	73,27
4,26	14,97	-1,35	-8,8	71,18
5,97	9,05	0,92	-0,2	70,15
2,63	15,43	5,96	4,9	117,33
0,62	12,43	17,28	-8,72	122,84
9,85	9,87	0,08	-13,62	33,66
5,32	10,33	2,06	4,1	77,31
8,93	10,08	-14,06	0,92	39,73
4,63	18,18	9,95	15,42	157,65
2,3	9,1	-2,26	-5,37	64,92
5,4	7,23	2,8	0,18	73,62
2,41	16,53	11,17	-7,28	112,9
1,98	8,67	-4,89	8,25	82,15
6,69	9,54	9,1	-1,14	79,1
8,57	13,52	1,89	0,74	90,78
7,96	6,92	5,75	-4,34	61,71
1,82	8,5	-4,35	-0,77	56,52
5,25	6,93	8,54	-7,38	65,96
4,83	9,7	3,79	-14,49	36,72
6,48	7,23	-0,86	-11,47	35,79
3,35	6,69	-11,41	0,68	38,77
2,26	8,87	-4,74	-0,92	74,21
6,62	11,33	-6,5	2,36	58,23
2,58	10,55	-0,37	1,26	85,93
-2,44	9,8	10,72	-4,07	106,77
5,37	14,52	-10,89	-7,56	42,01
-1,17	11,12	2,28	2,15	85,55
1,35	18,89	0,46	5,4	116,34
4,42	11,37	1,47	-5,84	67,07
9,41	8,71	-7,27	21,59	89,35
11,53	5,73	-2,38	-14,62	24,22
9,17	12,92	-10,96	-17,71	7,9
8,1	11,53	-7,1	-22,49	3,83
3,27	14,26	13,36	9,48	144,5
5,46	11,48	5,18	-7,5	66,46
6,68	14,35	-5,01	-8,41	53,39
5,95	9,78	1,05	-2,1	63,29
3,38	15,04	10,26	-19,95	68,01

0,57	17,3	-2,16	-5,03	80,75
8,56	9,3	6,56	-2,95	76,95
-0,24	5,82	0,14	0,16	58,98
2,88	7,89	3,4	-8,19	67,39
6,9	13,8	-2,88	-8,13	69,72
1,26	3,52	-1,33	-4,88	42,57
4,89	9,54	-3,63	-5,48	58,9
2,23	7,52	-5,13	2,68	62,13
2,71	7,27	0,77	7,96	95,52
4,19	12,39	2,84	-15	61,19
4,71	12,51	-0,27	0,78	87,87

Итоговый тест

1. Каким не может быть значение коэффициента корреляции

- A) нулевым
- B) отрицательным
- C) больше единицы

ANSWER: C

2. Что позволяет оценить коэффициент парной корреляции?

- A) Силу линейной связи
- B) Качественные различия признаков
- C) Мету изменчивости признака

ANSWER: A

3. Как проверяется статистическая гипотеза?

- A) Расчетом наблюдаемого значения критерия и сравнением с пороговым значением
- B) Перебором различных вариантов наблюдаемых значений
- C) Проведением повторных экспериментов

ANSWER: A

4. Что такое репрезентативная выборка?

- A) Выборка бесконечного размера
- B) Выборка, содержащая разные значения признака
- C) Выборка, по которой можно судить о генеральной совокупности

ANSWER: C

5. Что представляет собой нейронная сеть?

- A) Цепь отдельных элементов-нейронов
- B) Сеть искусственных нейронов
- C) Сеть передачи данных

ANSWER: B

5 ИНДИКАТОРЫ СФОРМИРОВАННЫХ КОМПЕТЕНЦИЙ ВЫПУСКНИКА ПРОГРАММЫ (при необходимости)

В результате освоения программы у слушателя сформированы компетенции:

Таблица 5

Компетенция (код, содержание)	Индикаторы
<p>А-6. Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры</p>	<ol style="list-style-type: none"> 1. Использует в своей работе понятийный аппарат из сферы работы с большими данными. 2. Демонстрирует умения извлечения, проверки и очистки данных построения и применения инструментальных средств обработки больших данных. 3. Владеет современными инструментальными средствами, технологической инфраструктурой анализа больших данных
<p>УК -1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</p>	<p>Демонстрирует способность работать с большими данными, осуществлять их подготовку данных и проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика,</p>