

МИНОБРНАУКИ РОССИИ
**Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Институт компьютерных наук и технологий

Авторы-составители: **Городилов Алексей Юрьевич**

Рабочая программа дисциплины
ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ
Код УМК 100508

Утверждено
Протокол №1
от «28» июня 2024 г.

Пермь, 2024

1. Наименование дисциплины

Введение в анализ данных

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление подготовки: **01.03.02** Прикладная математика и информатика
направленность Инженерия программного обеспечения

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Введение в анализ данных** у обучающегося должны быть сформированы следующие компетенции:

01.03.02 Прикладная математика и информатика (направленность : Инженерия программного обеспечения)

ОПК.1 Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности

Индикаторы

ОПК.1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические и физические объекты

4. Объем и содержание дисциплины

Направление подготовки	01.03.02 Прикладная математика и информатика (направленность: Инженерия программного обеспечения)
форма обучения	очная
№№ семестров, выделенных для изучения дисциплины	5
Объем дисциплины (з.е.)	4
Объем дисциплины (ак.час.)	144
Контактная работа с преподавателем (ак.час.), в том числе:	68
Проведение лекционных занятий	34
Проведение практических занятий, семинаров	34
Самостоятельная работа (ак.час.)	76
Формы текущего контроля	Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (2)
Формы промежуточной аттестации	Экзамен (5 семестр)

5. Аннотированное описание содержания разделов и тем дисциплины

Основы языка программирования Python

Возможности языка Python. Среда разработки. Основы синтаксиса Python. Основные типы данных. Алгоритмические конструкции: условный оператор и его варианты; программирование циклических алгоритмов.

Стандартные и пользовательские функции.

Кортежи. Списки и базовые алгоритмы их обработки. Ввод-вывод данных из файлов

Обработка строковых данных. Работа со словарями

Структуры данных: создание, свойства, вывод значений, срезы данных / копирование и ссылки.

Основные операции: удаление / фильтрация по условию, добавление / выравнивание / сортировка данных, работа с индексами, чистка данных.

Математические операции. Операции объединения данных, агрегирование данных. Первичный статистический анализ данных

Основные типы графиков и диаграмм и их реализация с помощью библиотек Matplotlib и Seaborn.

Математико-статистический анализ данных

Корреляционный и регрессионный анализ

Цели корреляционного и регрессионного анализа. Двумерная и трехмерная корреляционная модели.

Методы оценки корреляционных моделей.

Двумерная линейная регрессионная модель. Корреляционный и регрессионный анализ.

Понятие факторного анализа

Введение в машинное обучение. Обучение с учителем.

Сущность технологии Data Mining как процесса обнаружения в сырых данных практически полезных и доступных интерпретации и использованию знаний. Отличия Data Mining от других методов анализа данных.

Задачи Data Mining: классификация, прогнозирование, кластеризация, поиск ассоциаций, визуализация.

Примеры применения Data Mining в банковском деле, страховании, маркетинге, розничной торговле, анализе производственных ситуаций, в медицине, научных исследованиях и т.д.

Методы классификации и прогнозирования

Дерево решений как способ представления правил в иерархической последовательной структуре.

Обучение модели, созданной на основе дерева решений. Процесс конструирования дерева решений.

Метод опорных векторов (Support Vector Machine – SVM). Линейный SVM. Метод «ближайшего соседа» и метод «рассуждения по аналогии». Метод k-ближайших соседей для решения задач классификации и прогнозирования.

Наивный Байесовский классификатор.

Кластерный анализ и поиск ассоциативных правил

Методы кластерного анализа: иерархические и итеративные методы.

Кластерный анализ в задачах Data Mining. Иерархические методы кластерного анализа.

Итеративные методы кластерного анализа. Алгоритм k-средних.

Ассоциативные правила. Методы поиска ассоциативных правил.

Практическое применение технологий обработки данных

Примеры задач, связанных с анализом данных.

Развертывание систем аналитической обработки данных.

Хранилища данных.

Решение индивидуальных задач анализа данных и машинного обучения.

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторные занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Пальмов, С. В. Интеллектуальный анализ данных : учебное пособие / С. В. Пальмов. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт].
<http://www.iprbookshop.ru/75376.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. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт].
<http://www.iprbookshop.ru/26445>
3. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2019. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. <https://www.urait.ru/bcode/432851>
4. 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. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт].
<http://www.iprbookshop.ru/26444>

Дополнительная:

1. Буйначев, С. К. Основы программирования на языке Python : учебное пособие / С. К. Буйначев, Н. Ю. Боклаг ; под редакцией Ю. В. Песин. — Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2014. — 92 с. — ISBN 978-5-7996-1198-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <http://www.iprbookshop.ru/66183.html>
2. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2019. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт].
<https://www.urait.ru/bcode/432178>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<http://www.psu.ru/elektronnye-resursy-dlya-psu> Электронные ресурсы для ПГНИУ

<http://window.edu.ru/> Единое окно доступа к образовательным ресурсам

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Введение в анализ данных** предполагает использование следующего программного обеспечения и информационных справочных систем:

доступ в режиме on-line в Электронную библиотечную систему (ЭБС);

- доступ в электронную информационно-образовательной среду университета.

Необходимое лицензионное и (или) свободно распространяемое программное обеспечение:

- приложение позволяющее просматривать и воспроизводить медиаконтент PDF-файлов «Adobe Acrobat Reader DC»;
- офисный пакет приложений «LibreOffice», Alt Linux;
- пакета Anaconda для Python версии 3.

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для лекционных занятий требуется аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения практических занятий - компьютерный класс. Состав оборудования определен в Паспорте компьютерного класса (требуется доступ в Интернет и должен быть установлен пакет Anaconda for Python 3)

Для групповых (индивидуальных) консультаций - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения текущего контроля - аудитория, оснащенная меловой (и) или маркерной доской.

Самостоятельная работа студентов: аудитория, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет», с обеспеченным доступом в электронную информационно-образовательную среду университета, помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.
5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет LibreOffice.

Справочно-правовая система «КонсультантПлюс»

Фонды оценочных средств для аттестации по дисциплине
Введение в анализ данных

Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания

ОПК.1

Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
ОПК.1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические и физические объекты	<p>ОПК.1.2 Знать:</p> <ul style="list-style-type: none">- основные элементы языка программирования Python;- специальные структуры данных и библиотеки Python для обработки и анализа данных;- понятие больших данных, инструменты для работы с большими данными;- основные понятия машинного обучения;- математико-статистические методы анализа данных, такие как корреляционный анализ, регрессионный анализ, факторный анализ;- методы классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети;- методы кластерного анализа;- методы поиска ассоциативных правил. <p>Уметь:</p> <ul style="list-style-type: none">- использовать основные алгоритмические конструкции, типы и структуры данных языка Python;- применять библиотеки numpy и scipy для работы с числовой информацией;- применять библиотеку pandas для анализа табличных данных;- применять библиотеки matplotlib и seaborn для визуализации при обработке данных;- строить линейные модели в задачах регрессии и классификации;- строить решающие деревья и леса решающих деревьев.	<p>Неудовлетворител</p> <p>Отсутствие знаний или фрагментарные знания:</p> <ul style="list-style-type: none">- основных элементов языка программирования Python;- специальных структур данных и библиотек Python для обработки и анализа данных;- понятия больших данных, инструментов для работы с большими данными;- основных понятий машинного обучения;- математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ;- методов классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети;- методов кластерного анализа;- методов поиска ассоциативных правил. <p>Не умеет:</p> <ul style="list-style-type: none">- использовать основные алгоритмические конструкции, типы и структуры данных языка Python;- применять библиотеки numpy и scipy для работы с числовой информацией;- применять библиотеку pandas для анализа табличных данных;- применять библиотеки matplotlib и seaborn для визуализации при обработке данных;- строить линейные модели в задачах регрессии и классификации;- строить решающие деревья и леса решающих деревьев. <p>Не владеет навыками применения библиотек</p>

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
	<p>для анализа табличных данных;</p> <ul style="list-style-type: none"> - применять библиотеки matplotlib и seaborn для визуализации при обработке данных; - строить линейные модели в задачах регрессии и классификации; - строить решающие деревья и леса решающих деревьев. <p>Владеть навыками применения библиотек анализа данных для решения прикладных задач.</p>	<p>Неудовлетворител анализа данных для решения прикладных задач.</p> <p>Удовлетворительн Сформированы базовые знания:</p> <ul style="list-style-type: none"> - основных элементов языка программирования Python; - специальных структур данных и библиотек Python для обработки и анализа данных; - понятия больших данных, инструментов для работы с большими данными; - основных понятий машинного обучения; - математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ; - методов классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети; - методов кластерного анализа; - методов поиска ассоциативных правил. <p>В целом умеет, допуская некритичные ошибки:</p> <ul style="list-style-type: none"> - использовать основные алгоритмические конструкции, типы и структуры данных языка Python; - применять библиотеки numpy и scipy для работы с числовой информацией; - применять библиотеку pandas для анализа табличных данных; - применять библиотеки matplotlib и seaborn для визуализации при обработке данных; - строить линейные модели в задачах регрессии и классификации; - строить решающие деревья и леса решающих деревьев. <p>Владеет базовыми навыками применения библиотек анализа данных для решения прикладных задач.</p> <p>Хорошо Сформированы увереные, но не глубокие знания:</p> <ul style="list-style-type: none"> - основных элементов языка

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p style="text-align: center;">Хорошо</p> <p>программирования Python;</p> <ul style="list-style-type: none"> - специальных структур данных и библиотек Python для обработки и анализа данных; - понятия больших данных, инструментов для работы с большими данными; - основных понятий машинного обучения; - математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ; - методов классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети; - методов кластерного анализа; - методов поиска ассоциативных правил. <p>Умеет:</p> <ul style="list-style-type: none"> - использовать основные алгоритмические конструкции, типы и структуры данных языка Python; - применять библиотеки numpy и scipy для работы с числовой информацией; - применять библиотеку pandas для анализа табличных данных; - применять библиотеки matplotlib и seaborn для визуализации при обработке данных; - строить линейные модели в задачах регрессии и классификации; - строить решающие деревья и леса решающих деревьев. <p>Владеет основными навыками применения библиотек анализа данных для решения прикладных задач.</p> <p style="text-align: center;">Отлично</p> <p>Сформированы систематические глубокие знания:</p> <ul style="list-style-type: none"> - основных элементов языка программирования Python; - специальных структур данных и библиотек Python для обработки и анализа данных; - понятия больших данных, инструментов для работы с большими данными; - основных понятий машинного обучения;

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p style="text-align: center;">Отлично</p> <p>- математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ;</p> <p>- методов классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети;</p> <p>- методов кластерного анализа;</p> <p>- методов поиска ассоциативных правил.</p> <p>В совершенстве умеет:</p> <p>- использовать основные алгоритмические конструкции, типы и структуры данных языка Python;</p> <p>- применять библиотеки numpy и scipy для работы с числовой информацией;</p> <p>- применять библиотеку pandas для анализа табличных данных;</p> <p>- применять библиотеки matplotlib и seaborn для визуализации при обработке данных;</p> <p>- строить линейные модели в задачах регрессии и классификации;</p> <p>- строить решающие деревья и леса решающих деревьев.</p> <p>Владеет устойчивыми навыками применения библиотек анализа данных для решения прикладных задач.</p>

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 50 до 60

«неудовлетворительно» / «незачтено» менее 50 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические и физические объекты	Основы языка программирования Python Письменное контрольное мероприятие	Знания:- основных элементов языка программирования Python; - специальных структур данных и библиотек Python для обработки и анализа данных;- математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ. Умение:- использовать основные алгоритмические конструкции, типы и структуры данных языка Python; - применять библиотеки numpy и scipy для работы с числовой информацией;- применять библиотеку pandas для анализа табличных данных;- применять библиотеки matplotlib и seaborn для визуализации при обработке данных.

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические и физические объекты	Введение в машинное обучение. Обучение с учителем. Письменное контрольное мероприятие	Знания:- понятия больших данных, инструментов для работы с большими данными;- основных понятий машинного обучения;- математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ;- методов классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети;- методов кластерного анализа. Умение:- строить линейные модели в задачах регрессии и классификации;- строить решающие деревья и леса решающих деревьев.
ОПК.1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические и физические объекты	Практическое применение технологий обработки данных Итоговое контрольное мероприятие	Знания:- математико-статистических методов анализа данных, таких как корреляционный анализ, регрессионный анализ, факторный анализ;- методов классификации и прогнозирования: регрессия, деревья решений, метод опорных векторов, байесовская классификация, метод ближайших соседей, нейронные сети;- методов поиска ассоциативных правил. Умение:- применять библиотеку pandas для анализа табличных данных;- применять библиотеки matplotlib и seaborn для визуализации при обработке данных;- строить линейные модели в задачах регрессии и классификации;- строить решающие деревья и леса решающих деревьев. Владение навыками применения библиотек анализа данных для решения прикладных задач.

Спецификация мероприятий текущего контроля

Основы языка программирования Python

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **15**

Показатели оценивания	Баллы
Решены задачи на умение применять библиотеку pandas для анализа табличных данных, работу с объектами Series и DataFrame	9
Решены задачи на знание и умение использовать основные элементы языка программирования Python, основные алгоритмические конструкции, типы и структуры данных языка Python	7
Решены задачи на работу с NumPy-массивами и сериями, умение применять библиотеки numpy и scipy для работы с числовой информацией	6
Решены задачи на методы работы со списками и строками, форматирование строк.	4
Решены задачи на методы работы со множествами, кортежами, словарями.	4

Введение в машинное обучение. Обучение с учителем.

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **15**

Показатели оценивания	Баллы
Решены задачи на применение методов классификации и прогнозирования	8
Решены задачи на умение строить линейные модели в задачах регрессии и классификации	7
Решены задачи на применение методов кластерного анализа	5
Решены задачи на умение применять библиотеки matplotlib и seaborn для визуализации при обработке данных	5
Решены задачи на умение строить решающие деревья и леса решающих деревьев	5

Практическое применение технологий обработки данных

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **20**

Показатели оценивания	Баллы
Тест по теории. 10 вопросов тестового типа на знание основных определений и понятий анализа данных и машинного обучения	10
Выполнение индивидуального задания на анализ данных. При решении использованы библиотеки языка Python.	8
Выполнение индивидуального задания на анализ данных. Использованы библиотеки matplotlib или seaborn для визуализации.	7
Выполнение индивидуального задания на анализ данных. Выполнен поиск ассоциативных	5

правил.	
Выполнение индивидуального задания на анализ данных. В работе построено решающее дерево.	5
Выполнение индивидуального задания на анализ данных. В работе решена задача классификации, кластеризации или прогнозирования.	5