

**ГАОУ ВО «ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ НАРОДНОГО ХОЗЯЙСТВА»**

*Утверждены решением
Ученого совета ДГУНХ,
протокол № 11
от 06 июня 2023 г.*

**КАФЕДРА «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ»**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

ПО ДИСЦИПЛИНЕ

«ТЕХНОЛОГИИ АНАЛИЗА БОЛЬШИХ ДАННЫХ»

**Направление подготовки
38.04.05 Бизнес-информатика,
профиль «УПРАВЛЕНИЕ IT –ПРОЕКТАМИ И
ПРОДУКТАМИ»**

Уровень высшего образования – магистратура

УДК 681.3 (076)
ББК 32.97 я 7 П 13

Составитель – Кобзаренко Дмитрий Николаевич, доктор технических наук, профессор кафедры «Информационные технологии и информационная безопасность» ДГУНХ.

Внутренний рецензент – Савина Елена Владимировна, кандидат физико-математических наук, доцент кафедры «Информационные технологии и информационная безопасность» ДГУНХ.

Внешний рецензент – Абдурагимов Гусейн Эльдарханович, кандидат физико-математических наук, доцент кафедры "Математические методы в экономике" Дагестанского государственного университета.

Представитель работодателя – Ботвин Тимур Анатольевич, руководитель международных запусков Яндекс.Маркет ООО «Яндекс.Маркет».

Оценочные материалы по дисциплине «Технологии анализа больших данных» разработаны в соответствии с требованиями федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.04.05 Бизнес-информатика, утвержденного приказом Министерства образования и науки Российской Федерации от 12 августа 2020 г., № 990, в соответствии с приказом Министерства науки и высшего образования от 6.04.2021 г., № 245 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры».

Оценочные материалы по дисциплине «Технологии анализа больших данных» размещены на официальном сайте www.dgunh.ru

Кобзаренко Д.Н. Оценочные материалы по дисциплине «Технологии анализа больших данных» для направления подготовки 38.04.05 Бизнес-информатика, профиль «Управление IT –проектами и продуктами». – Махачкала: ДГУНХ, 2023 г. – 29 с.

Рекомендованы к утверждению Учебно-методическим советом ДГУНХ 05 июня 2023 г.

Рекомендованы к утверждению руководителем основной профессиональной образовательной программы высшего образования – программы магистратуры по направлению подготовки 38.04.05 Бизнес-информатика, профиль «Управление IT –проектами и продуктами», к.пед.н., Гасановой З.А.

Одобрены на заседании кафедры «Информационные технологии и информационная безопасность» 31 мая 2023 г., протокол № 10.

Содержание

НАЗНАЧЕНИЕ ОЦЕНОЧНЫХ МАТЕРИАЛОВ.....	4
РАЗДЕЛ 1. Перечень компетенций с указанием видов оценочных средств в процессе освоения дисциплины.....	5
1.1 Перечень формируемых компетенций.....	5
1.2 Перечень компетенций с указанием видов оценочных средств.....	5
РАЗДЕЛ 2. Задания, необходимые для оценки планируемых результатов обучения по дисциплине.....	8
РАЗДЕЛ 3. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания.....	19
РАЗДЕЛ 4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков, характеризующие этапы формирования компетенций.....	23
Лист актуализации оценочных материалов по дисциплине.....	29

НАЗНАЧЕНИЕ ОЦЕНОЧНЫХ МАТЕРИАЛОВ

Оценочные материалы для текущего контроля успеваемости (оценивания хода освоения дисциплин), для проведения промежуточной аттестации (оценивания промежуточных и окончательных результатов обучения по дисциплине) обучающихся по дисциплине «Технологии анализа больших данных» на соответствие их учебных достижений поэтапным требованиям образовательной программы высшего образования 38.04.05 Бизнес-информатика, профиль «Управление IT – проектами и продуктами».

Оценочные материалы по дисциплине «Технологии анализа больших данных» включают в себя: перечень компетенций с указанием видов оценочных средств в процессе освоения дисциплины; описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания; типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения ОПОП; методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Оценочные материалы сформированы на основе ключевых принципов оценивания:

- валидности: объекты оценки должны соответствовать поставленным целям обучения;
- надежности: использование единообразных стандартов и критериев для оценивания достижений;
- объективности: разные обучающиеся должны иметь равные возможности для достижения успеха.

Основными параметрами и свойствами оценочных материалов являются:

- предметная направленность (соответствие предмету изучения конкретной дисциплины);
- содержание (состав и взаимосвязь структурных единиц, образующих содержание теоретической и практической составляющих дисциплины);
- объем (количественный состав оценочных материалов);
- качество оценочных материалов в целом, обеспечивающее получение объективных и достоверных результатов при проведении контроля с различными целями.

РАЗДЕЛ 1. Перечень компетенций с указанием видов оценочных средств в процессе освоения дисциплины

1.1 Перечень формируемых компетенций

код компетенции	формулировка компетенции
ПК	ПРОФЕССИОНАЛЬНЫЕ КОМПЕТЕНЦИИ
ПК-1	Способен реализовывать проекты создания и развития ИТ-инфраструктуры предприятия

1.2 Перечень компетенций с указанием видов оценочных средств

<i>Формируемые компетенции</i>	<i>Код и наименование индикатора достижения компетенции</i>	<i>Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций</i>	<i>Уровни освоения компетенции</i>	<i>Критерии оценивания сформированности компетенций</i>	<i>Виды оценочных средств</i>
ПК-1. Способен реализовывать проекты создания и развития ИТ-инфраструктуры предприятия.	ПК-1.2. Совершает ИТ-инфраструктуру предприятия с использованием современных цифровых решений	<u>Знать:</u> – основные понятия термина Большие Данные; – основные понятия Data Mining; – основные способы хранения и анализа данных; – методы извлечения знаний из данных (кластеризация,	Пороговый уровень	Обучающийся слабо (частично) знает основные понятия термина Большие Данные, основные понятия Data Mining, основные способы хранения и анализа данных, методы извлечения знаний из данных.	Блок А – задания репродуктивного уровня - тестовые задания; - вопросы для обсуждения (устный опрос).
			Базовый уровень	Обучающийся с незначительными ошибками и отдельными пробелами знает основные понятия термина Большие Данные, основные понятия Data Mining, основные способы хранения и анализа данных, методы извлечения знаний из данных.	

		классификация, прогнозирование, регрессия, нейронные сети).	Продвинутый уровень	Обучающийся с требуемой степенью полноты и точности знает основные понятия термина Большие Данные, основные понятия Data Mining, основные способы хранения и анализа данных, методы извлечения знаний из данных.	
		<p>Уметь:</p> <ul style="list-style-type: none"> – анализировать современные потоки данных; – находить, извлекать и структурировать данные; – работать с программными средствами для хранения и анализа данных; – разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия 	Пороговый уровень	Обучающийся слабо (частично) умеет анализировать современные потоки данных; находить, извлекать и структурировать данные; работать с программными средствами для хранения и анализа данных; разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия.	Блок В – задания для реконструктивного уровня - тестовые задания; - рефераты.
			Базовый уровень	Обучающийся с незначительными затруднениями умеет анализировать современные потоки данных; находить, извлекать и структурировать данные; работать с программными средствами для хранения и анализа данных; разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия.	

		.	Продвинутый уровень	Обучающийся умеет анализировать современные потоки данных; находить, извлекать и структурировать данные; работать с программными средствами для хранения и анализа данных; разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия.	
		<p>Владеть:</p> <ul style="list-style-type: none"> – терминологией Больших данных – терминологией Data Mining; – основными способами хранения и анализа данных; – навыками извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети). 	Пороговый уровень	Обучающийся слабо (частично) владеет терминологией Больших данных, терминологией Data Mining, основными способами хранения и анализа данных, навыками извлечения знаний из данных.	<p>Блок С – задания для практико-ориентированного уровня - лабораторные задания.</p>
	Базовый уровень		Обучающийся с незначительными затруднениями владеет терминологией Больших данных, терминологией Data Mining, основными способами хранения и анализа данных, навыками извлечения знаний из данных		
	Продвинутый уровень		Обучающийся свободно владеет терминологией Больших данных, терминологией Data Mining, основными способами хранения и анализа данных, навыками извлечения знаний из данных.		

РАЗДЕЛ 2. Задания, необходимые для оценки планируемых результатов обучения по дисциплине

Для проверки сформированной компетенции

ПК-1. Способен реализовывать проекты создания и развития ИТ-инфраструктуры предприятия

ПК-1.2. Совершенствует ИТ-инфраструктуру предприятия с использованием современных цифровых решений

Блок А. Задания репродуктивного уровня («знать»)

А.1 Тестовые задания

1. Суммарная оперативная память IBM Watson составляет порядка:

- А) 100 гигабайт
- Б) 5000 терабайт
- В) 10 зетабайт
- Г) **15 терабайт**

2. Кто ввел термин Большие данные?

- А) **Клиффорд Линч**
- Б) Алан Тьюринг
- В) Бьерн Страуструп
- Г) Дональд Кнут

3. Какие данные занимают больше мировой памяти относительно остальных?

- А) Structured Data
- Б) **Unstructured Data**
- В) Semi-Structured Data
- Г) Quasi-Structured Data

4. BigData – это ...

- А) Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации, или обработки.
- Б) **Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.**
- В) Колоссальный объем данных, собранных человечеством.
- Г) Класс в Java, предназначенный для хранения данных от 100 Гб

5. Какая компания создала технологию MapReduce?

- А) **Google**
- Б) Yahoo
- В) EMC
- Г) Oracle

6. Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:

- А) Структурированными
- Б) Полуструктурированными**
- В) Квазиструктурированными
- Г) Неструктурированными

7. Что означает термин «Big Data» в информационных технологиях?

- А) Комплексный набор методов для создания файлов большого объема
- Б) Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.**
- В) Файлы с большим количеством данных.
- Г) Представление времени, дня, месяца и года в качестве значения количества миллисекунд, прошедших с начала нашей эры.

8. Данные имеющие определенный тип, формат и структуру (например, транзакционные данные) являются:

- А) Структурированными**
- Б) Полуструктурированными
- В) Квазиструктурированными
- Г) Неструктурированными

9. Чему примерно равен объем всей существующей на земле информации (в байтах)?

- А) 10^{11}
- Б) 10^{21}**
- В) $10^{1010101}$
- Г) 10^{171}

10. В каком году впервые был введен термин Большие данные?

- А) 2002
- Б) 2004
- В) 2006
- Г) 2008**

11. Что является средством анализа в BI?

- А) Карты показателей;**
- Б) Совместная работа и управление рабочими процессами;
- В) Информационные панели;
- Г) BI инфраструктура.

12. Основное умение исследователя данных?

- А) Умение находить наиболее важные элементы в хранимой информации
- Б) Уметь прогнозировать исход работы системы
- В) Находить скрытые логические связи в системе собранной информации**

Г) Отличать неструктурированные данные от структурированных

13. Что означает термин «Business Intelligence» в информационных технологиях?

А) Комплексный набор методов для создания бизнес планов.

Б) Методы и инструменты для перевода необработанной информации в осмысленную, удобную для восприятия форму.

В) Файлы, содержащие информацию о бизнес плане.

Г) Технологии, направленные на развитие бизнеса.

14. Языком, на котором был разработан RabbitMQ, является:

А) Java

Б) Python

В) C++

Г) Erlang

15. Что является главным результатом процесса Business Intelligence?

А) Возможность принятия решений для бизнеса

Б) Результаты интеллектуального анализа данных

В) Возможность использования искусственного интеллекта

Г) Получение структуризации данных после выполнения всех шагов процесса

16. Что из перечисленного не является средством анализа?

А) Продвинутая визуализация

Б) Reporting

В) Predictive Modelling

Г) Data Mining

17. Что относится к средствам предоставления информации в «Business Intelligence»?

А) Генератор нерегламентированных запросов

Б) Совместная работа и управление рабочими процессами

В) Предиктивное моделирование и Data Mining

Г) Карты показателей

18. Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является:

А) OLAP

Б) Data Mining

В) Predictive Modelling

Г) Data Science

19. Что не является целью процесса Business Intelligence?

А) Интерпретация большого количества данных;

Б) Моделирование исходов различных вариантов действий;

В) Модификация существующего программного обеспечения;

Г) Отслеживание результатов решений.

20. Что из этого не является реализацией Hadoop?

А) Google MapReduce

Б) Phoenix

В) GreenMint

Г) Qizmt

21. Какие из перечисленных пунктов являются достоинствами MapReduce?

А) Оптимальная производительность

Б) Эффективное применение в маленьких кластерах с небольшим объемом данных

В) Масштабируемость

Г) Отказоустойчивость

22. Какие компоненты являются частями MapReduce?

А) Task Tracker

Б) Name Node и Data Node

В) Job Tracker и Task Tracker

Г) Job Tracker, Task Tracker, Name Node и Data Node

23. Что такое Spark?

А) Инструмент для кластерных вычислений

Б) Графический движок

В) Библиотека для работы с графами

Г) Технология распределенных вычислений

24. Дайте определение Map Reduce...

А) Модель распределенных вычислений, предназначенная для параллельных вычислений над очень большими (до нескольких петабайт) объемами данных

Б) Набор компонентов и интерфейсов для распределенных файловых систем и общего ввода-вывода

В) Распределенная файловая система, работающая на больших кластерах типовых машин

Г) Распределенный сервис для коллекционирования, сбора, и перемещения больших массивов данных

25. Что из этого является недостатком MapReduce?

А) Фиксированный алгоритм обработки данных

Б) Масштабируемость

В) Отказоустойчивость

Г) Возможность автоматического распараллеливания

26. Какое API было добавлено в Hadoop v2.0?

- А) YAWN
- Б) YARN**
- В) SARN
- Г) DARN

27. Какая цель у NameNode в HDFS?

- А) Хранить индекс того, какая часть данных находится в каком узле**
- Б) Хранить имя файла, хранящегося в конкретном узле
- В) Хранить индекс узла, в котором хранится имя файла
- Г) Хранить имена узлов

28. Вертикальное масштабирование...

- А) Требует изменений в прикладных программах, работающих на таких системах
- Б) Не требует никаких изменений в прикладных программах, работающих на таких системах**
- В) Уменьшает производительность каждого компонента БД
- Г) Увеличивает скорость загрузки данных

29. Что, согласно теореме CAP (теореме Брюера), возможно обеспечить в любой реализации распределённых вычислений?

- А) Только согласованность данных
- Б) Только доступность данных
- В) Согласованность данных, доступность данных, устойчивость к разделению
- Г) Не более двух свойств из трёх вышеприведённых**

30. Что означает термин NoSQL?

- А) Не SQL
- Б) Не только SQL**
- В) Без SQL
- Г) SQL – плохо

31. Разбиение системы на более мелкие структурные компоненты и разнесение их по отдельным физическим машинам (или их группам), и (или) увеличение количества серверов, параллельно выполняющих одну и ту же функцию, это:

- А) Горизонтальное масштабирование**
- Б) Вертикальное масштабирование
- В) Master- slave репликация
- Г) Peer-to-peer репликация

32. Выделение таблицы или группы таблиц на отдельный сервер это...

- А) Горизонтальное масштабирование
- Б) Вертикальное масштабирование
- В) Горизонтальный шардинг
- Г) Вертикальный шардинг**

33. Преподнесение какой-либо полезной информации в форме интересного рассказа – это...

- А) Сторителлинг
- Б) Инфографика
- В) Бизнес аналитика
- Г) Картограмма

34. Основная цель статистического анализа:

- А) Поиск генеральной совокупности
- Б) Выяснение свойств генеральной совокупности**
- В) Сравнение генеральных совокупностей
- Г) Выявление последовательности входного набора

35. Определённое предположение о распределении вероятностей, лежащем в основе

наблюдаемой выборки данных, - это...

- А) Статистический критерий
- Б) Статистическая выборка
- В) Статистическая гипотеза**
- Г) Задача кластеризации

36. С помощью какого алгоритма можно найти ассоциативное правило?

- А) Алгоритм apriori**
- Б) Алгоритм k-means
- В) Алгоритм c-means
- Г) Иерархический алгоритм

A.2 Вопросы для обсуждения

1. Что означает термин «Big Data» в информационных технологиях?
2. Что является основной целью обработки Big Data?
3. Кто и в каком году впервые ввел термин «Big Data»?
4. Какие главные характеристики Big Data?
5. Какие данные занимают больше мировой памяти относительно остальных?
6. Какие понятия содержит в себе принцип трех "V"?
7. С какого года Большие данные изучаются как академический предмет в вузовских программах по науке о данных?
8. Что является примером квази-структурированных данных?
9. Как назывался первый суперкомпьютер, оснащенный вопросно-ответной системой искусственного интеллекта?
10. Чем характеризуются "Большие данные"?
11. Что является главным результатом процесса Business Intelligence?
12. Что означает термин «Business Intelligence» в информационных технологиях?
13. Расшифруйте аббревиатуру OLAP.
14. Что относится к средствам предоставления информации в Business

Intelligence?

15. Что относится к средствам интеграции в «Business Intelligence»?
16. Какие цели ставит перед собой Data Science?
17. Что такое жизненный цикл аналитики данных?
18. Дайте определение термину «предиктивное моделирование»?
19. Что такое ETL?
20. Какова роль BI-аналитика в проекте?
21. В чем преимущества решений на базе Hadoop?
22. Что такое MapReduce, какими достоинствами и недостатками обладает MapReduce?
23. Какому основному принципу следует HDFS и какой размер блока по умолчанию в HDFS?
24. Какие функции выполняет NameNode в HDFS?
25. Какой узел отвечает за репликацию данных в Hadoop?
26. Что является отличительной особенностью NoSQL?
27. Что, согласно теореме CAP, возможно обеспечить в любой реализации распределённых вычислений?
28. Что относится к плюсам репликации?
29. Что относится к преимуществам нереляционных БД?
30. Что такое шардинг?
31. Какие достоинства у Amazon S3?
32. В чем особенности хранения в Amazon S3?
33. Что такое дедупликация данных?
34. Какие типы визуализации можно выделить?
35. Какие основные типы Data Mining?
36. Какие категории Web Mining можно выделить?
37. Что изучает статистика?
38. К каким алгоритмам классификации относится метод ближайших соседей?
39. Что является целью кластеризации?
40. Язык Python и библиотеки для анализа данных.

Блок В. Задания реконструктивного уровня («уметь»)

В.1 Тестовые задания

Задача №1. Что отобразится на экране после исполнения кода на Python?

```
s = ['Махачкала', 'Дербент', 'Кизляр', 'Касумкент']
s.sort()
d = {}
i = 1
for x in s:
    i += 2
    d[x] = i + 1
k = 0
```

```
for i in range(d['Дербент'] * d['Махачкала'] % 35):
    if i != 2:
        k += i
print(k)
```

Ответ: 8

Задача №2. Что отобразится на экране после исполнения кода на Python?

```
import numpy as np
a = np.array(list([x for x in range(10)]))
b = np.arange(10)
c = a[2:] - b[:-2]
print(sum(c) + max(c) + min(c))
```

Ответ: 20

Задача №3. Что отобразится на экране после исполнения кода на Python?

```
import numpy as np
a = np.array([[1, 2, 3],
              [4, 5, 6],
              [7, 8, 9]])
a += 1
print(sum(a[1]) + max(a[0]) - min(a[2]))
```

Ответ: 14

Задача №4. Что отобразится на экране после исполнения кода на Python?

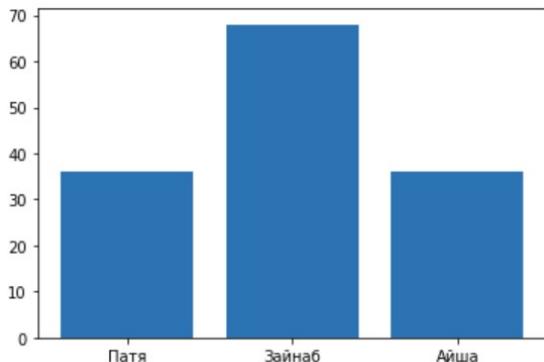
```
import pandas as pd
d = {'names': ['Патя', 'Зурик', 'Мага',
              'Зайнаб', 'Аташка', 'Айша'],
     'sex': ['Ж', 'М', 'М', 'Ж', 'М', 'Ж'],
     'age': [18, 22, 26, 34, 25, 18],
     'height': [170, 168, 172, 165, 167, 160],
     'merried': [True, False, True, True, False, False]}
data = pd.DataFrame(d)
my_list = list(data[(data['merried'] != True) &
                   (data['age'] < 25) &
                   (data['height'] <= 170)].names)
print(my_list[0], '+', my_list[1], '= Love')
```

Ответ: Зурик + Айша = Love

Задача №5. Какая из строк отрисовки графика в программном коде на Python соответствует построенному графику?

```
import pandas as pd
import matplotlib.pyplot as plt
d = {'names': ['Патя', 'Зурик', 'Мага',
              'Зайнаб', 'Аташка', 'Айша'],
     'sex': ['Ж', 'М', 'М', 'Ж', 'М', 'Ж'],
```

```
'age': [18, 22, 26, 34, 25, 18],
data = pd.DataFrame(d)
# ОТРИСОВКА ГРАФИКА
plt.show()
```



- А) `plt.plot(data[data.sex!='M'].names, data[data.sex!='M'].age*2)`
 Б) `plt.plot(data[data.sex=='F'].names, data[data.sex=='F'].age+2)`
В) `plt.bar(data[data.sex!='M'].names, data[data.sex!='M'].age*2)`
 Г) `plt.bar(data[data.sex=='F'].names, data[data.sex=='F'].age+2)`
 Д) `plt.hist(data[data.sex!='M'].names, data[data.sex!='M'].age*2)`
 Е) `plt.hist(data[data.sex=='F'].names, data[data.sex=='F'].age+2)`

В.2 Рефераты (темы)

1. Современные тенденции развития Big Data в мировой компьютерной индустрии.
2. Современные тенденции развития Big Data в Российской Федерации.
3. Предпосылки развития Big Data в Республике Дагестан.
4. Современное состояние операционной системы Hadoop.
5. Отрасли народного хозяйства Республики Дагестан, в которых может быть задействована операционная система Hadoop.
6. Перечень народно-хозяйственных задач в Республике Дагестан, которые целесообразно решать с помощью технологии Map Reduce.
7. Применение технологий NoSQL в современных бизнес проектах.
8. Перспективы применения среды R для анализа статистических данных, необходимых для принятия решений в управлении городским округом Махачкала.
9. Отрасли народного хозяйства Республики Дагестан, в которых могут быть задействованы системы класса Data Mining.
10. Перспективы применения кластерного анализа и его результатов для управления экономикой Республики Дагестан.

Блок С. Задания практикоориентированного уровня для диагностирования сформированности компетенций («владеть»)

С.1 Лабораторные работы

Лабораторная работа №1. «Знакомство со средой Google Colaboratory и изучения основ языка Python»

Задание 1:

Ознакомиться с облачной средой Google Colaboratory и выполнить регистрацию аккаунта в google, если это необходимо.

Задание 2:

Изучить базовый набор синтаксиса языка Python, основные типы данных, служебные директивы, операторы условного перехода и циклов, списки, словари, кортежи, множества, функции, исключения.

Задание 3:

Решить предложенные задачи.

Лабораторная работа №2. «Работа с библиотекой NumPy»

Задание 1:

Изучить основные возможности библиотеки NumPy для работы с одномерными и многомерными массивами данных.

Задание 2:

Решить предложенные задачи.

Лабораторная работа №3. «Работа с библиотекой Pandas»

Задание 1:

Изучить основные возможности библиотеки Pandas для работы с дафреймами.

Задание 2:

Решить предложенные задачи.

Лабораторная работа №4. «Работа с библиотекой Matplotlib»

Задание 1:

Изучить основные возможности библиотеки Matplotlib для работы с графикой.

Задание 2:

Решить предложенные задачи.

Лабораторная работа №5. «Построение нейронной сети в библиотеке Keras»

Задание 1:

Изучить основы построения моделей нейронных сетей через библиотеку Keras.

Задание 2:

Построить нейронные сети по распознаванию рукописных цифр и геометрических фигур.

Блок Д. Задания для использования в рамках промежуточной аттестации

Д.1 Перечень вопросов на экзамен

1. Термин Big Data, цель и главные характеристики больших данных.
2. Принцип трех "V".
3. Структурированные, неструктурированные и квази-структурированные

данные.

4. Business Intelligence.
5. Цели и задачи Data Science.
6. Что такое жизненный цикл аналитики данных.
7. Технология Apache Hadoop.
8. Технология MapReduce, достоинства и недостатки.
9. Файловая система HDFS.
10. Хранилища NoSQL главные отличия от обычных БД.
11. Теорема CAP и ее следствия.
12. Amazon S3.
13. Основные задачи, требования и типы визуализации.
14. Data Mining и основные типы.
15. Классификация и кластеризация.
16. Язык Python и библиотеки для анализа данных.
17. Построение нейронных сетей на базе библиотеки Keras.

РАЗДЕЛ 3. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Балльно-рейтинговая система является базовой системой оценивания сформированности компетенций обучающихся очной формы обучения.

Итоговая оценка сформированности компетенции обучающихся в рамках балльно-рейтинговой системы осуществляется в ходе текущего контроля успеваемости, промежуточной аттестации и определяется как сумма баллов, полученных обучающимися в результате прохождения всех форм контроля.

Оценка сформированности компетенции по дисциплине складывается из двух составляющих:

✓ первая составляющая – оценка преподавателем сформированности компетенции в течение семестра в ходе текущего контроля успеваемости (максимум 100 баллов). Структура первой составляющей определяется технологической картой дисциплины, которая в начале семестра доводится до сведения обучающихся;

✓ вторая составляющая – оценка сформированности компетенции обучающихся на экзамене (максимум – 30 баллов).

Для студентов очно-заочной и заочной форм обучения применяется 4-балльная и бинарная шкалы оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся.

уровни освоения компетенций	продвинутый уровень	базовый уровень	пороговый уровень	допороговый уровень
100 – балльная шкала	85 и \geq	70 – 84	51 – 69	0 – 50
4 – балльная шкала	«отлично»	«хорошо»	«удовлетворительно»	«неудовлетворительно»

Шкала оценок при текущем контроле успеваемости по различным показателям

Показатели оценивания сформированности компетенций	Баллы	Оценка
Устный опрос	0-15	«неудовлетворительно» «удовлетворительно» «хорошо» «отлично»
Выполнение лабораторной работы	0-15	«неудовлетворительно» «удовлетворительно» «хорошо» «отлично»

Тестирование / контрольная работа	0-30	«неудовлетворительно» «удовлетворительно» «хорошо» «отлично»
Выполнение реферата	0-10	«неудовлетворительно» «удовлетворительно» «хорошо» «отлично»

**Соответствие критериев оценивания уровню освоения компетенций
по текущему контролю успеваемости**

Баллы	Оценка	Уровень освоения компетенций	Критерии оценивания
0-50	«неудовлетворительно»	допороговый уровень	Обучающийся не приобрел знания, умения и не владеет компетенциями в объеме, закрепленном рабочей программой дисциплины
51-69	«удовлетворительно»	пороговый уровень	Не менее 50% заданий, подлежащих текущему контролю успеваемости, выполнены без существенных ошибок
70-84	«хорошо»	базовый уровень	Обучающимся выполнено не менее 75% заданий, подлежащих текущему контролю успеваемости, или при выполнении всех заданий допущены незначительные ошибки; обучающийся показал владение навыками систематизации материала и применения его при решении практических заданий; задания выполнены без ошибок
85-100	«отлично»	продвинутый уровень	100% заданий, подлежащих текущему контролю успеваемости, выполнены самостоятельно и в требуемом объеме; обучающийся проявляет умение обобщать, систематизировать материал и применять его при решении практических заданий; задания выполнены с подробными пояснениями и

			аргументированными выводами
--	--	--	-----------------------------

Шкала оценок по промежуточной аттестации

<i>Наименование формы промежуточной аттестации</i>	<i>Баллы</i>	<i>Оценка</i>
Экзамен	0-30	«неудовлетворительно» «удовлетворительно» «хорошо» «отлично»

Соответствие критериев оценивания уровню освоения компетенций по промежуточной аттестации обучающихся

Баллы	Оценка	Уровень освоения компетенций	Критерии оценивания
0-9	«не зачтено»	допороговый уровень	Обучающийся не приобрел знания, умения и не владеет компетенциями в объеме, закрепленном рабочей программой дисциплины; обучающийся не смог ответить на вопросы
10-14	«зачтено»	пороговый уровень	Обучающийся дал неполные ответы на вопросы, с недостаточной аргументацией, практические задания выполнены не полностью, компетенции, осваиваемые в процессе изучения дисциплины сформированы не в полном объеме.
15-17	«зачтено»	базовый уровень	Обучающийся в целом приобрел знания и умения в рамках осваиваемых в процессе обучения по дисциплине компетенций; обучающийся ответил на все вопросы, точно дал определения и понятия, но затрудняется подтвердить теоретические положения практическими примерами; обучающийся показал хорошие знания по предмету, владение навыками систематизации

			материала и полностью выполнил практические задания
18-20	«зачтено»	продвинутый уровень	Обучающийся приобрел знания, умения и навыки в полном объеме, закрепленном рабочей программой дисциплины; терминологический аппарат использован правильно; ответы полные, обстоятельные, аргументированные, подтверждены конкретными примерами; обучающийся проявляет умение обобщать, систематизировать материал и выполняет практические задания с подробными пояснениями и аргументированными выводами

РАЗДЕЛ 4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков, характеризующие этапы формирования компетенций

Тестирование (контрольная работа) выполняется на подготовленных бланках, в которых имеется теоретическая часть – контрольные вопросы с вариантами ответов и практическая часть – задачи с полем ответа для записи результата. Для выполнения тестирования студенту отводится 90 минут времени. За правильные ответы на теоретические вопросы студент получает по 1-2 балла, за решенные задачи 3-5 баллов. Весь тест составляется из расчета максимального получения 30 баллов.

Методика оценивания выполнения тестов

Баллы	Оценка	Показатели	Критерии
25-30	«Отлично»	1. Полнота выполнения тестовых заданий; 2. Своевременность выполнения; 3. Правильность ответов на вопросы;	Выполнено более 85 % заданий предложенного теста, в заданиях открытого типа дан полный, развернутый ответ на поставленный вопрос
19-24	«Хорошо»	4. Самостоятельность тестирования.	Выполнено более 70 % заданий предложенного теста, в заданиях открытого типа дан полный, развернутый ответ на поставленный вопрос; однако были допущены неточности в определении понятий, терминов и др.
13-18	«Удовлетворительно»		Выполнено более 54 % заданий предложенного теста, в заданиях открытого типа дан неполный ответ на поставленный вопрос, в ответе не присутствуют доказательные примеры, текст со стилистическими и орфографическими ошибками.

0-11	«Неудовлетворительно»	Выполнено не более 53 % заданий предложенного теста, на поставленные вопросы ответ отсутствует или неполный, допущены существенные ошибки в теоретическом материале (терминах, понятиях).
------	-----------------------	---

Устный опрос проводится на практических занятиях по пройденному лекционному материалу. Студенту дается вопрос, на который предлагается дать развернутый ответ. Максимально за ответ на вопрос студент может получить 15 баллов.

Методика оценивания ответов на устные вопросы

Баллы	Оценка	Показатели	Критерии
13-15	«Отлично»	<ol style="list-style-type: none"> 1. Полнота данных ответов; 2. Аргументированность данных ответов; 3. Правильность ответов на вопросы. 	<p>Полно и аргументировано даны ответы по содержанию задания. Обнаружено понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только по учебнику, но и самостоятельно составленные. Изложение материала последовательно и правильно.</p>
10-12	«Хорошо»		<p>Студент дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1-2 ошибки, которые сам же исправляет.</p>

6-9	«Удовлетворительно»		<p>Студент обнаруживает знание и понимание основных положений данного задания, но:</p> <ol style="list-style-type: none"> 1) излагает материал неполно и допускает неточности в определении понятий или формулировке правил; 2) не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; 3) излагает материал непоследовательно и допускает ошибки.
0-5	«Неудовлетворительно»		<p>Студент обнаруживает незнание ответа на соответствующее задание, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал; отмечаются такие недостатки в подготовке студента, которые являются серьезным препятствием к успешному овладению последующим материалом.</p>

Лабораторные задания выполняются на лабораторных занятиях на основе методического руководства к выполнению лабораторной работы. Студент должен

выполнить задание к лабораторной работы, а затем ответить на контрольные вопросы, связанные с выполнением лабораторной работы. Максимально за выполнение лабораторной работы студент может получить 15 баллов.

Методика оценивания выполнения лабораторных заданий

Баллы	Оценка	Показатели	Критерии
13-15	«Отлично»	<ol style="list-style-type: none"> 1. Полнота выполнения лабораторной работы; 2. Своевременность выполнения лабораторной работы; 3. Правильность выполнения лабораторной работы. 	Выполнены все требования к лабораторной работе; разработана, отлажена и протестирована программа; даны правильные ответы на дополнительные вопросы.
10-12	«Хорошо»		Выполнены основные требования к лабораторной работе, имеются недочеты в разработке и тестировании программы; на дополнительные вопросы при защите даны неполные ответы.
6-9	«Удовлетворительно»		Требования к лабораторной работе выполнены не полностью, программа разработана, но в ней имеются существенные недостатки; допущены фактические ошибки при ответе на дополнительные вопросы.
0-5	«Неудовлетворительно»		Лабораторная работа не выполнена; обнаруживается существенное непонимание в ее выполнении.

Реферат выполняется студентом в течении всего срока аттестационного периода. Выполнив реферат, студент сдает его на проверку преподавателю. Если

реферат соответствует требованиям задания, то назначается его защита. На защите студент излагает суть реферата и отвечает на вопросы преподавателя относительно выполненной работы. Максимально за выполнение реферата студент может получить 10 баллов.

Методика оценивания выполнения рефератов

Баллы	Оценка	Показатели	Критерии
9-10	«Отлично»	1. Полнота выполнения реферата; 2. Своевременность выполнения реферата; 3. Правильность выполнения реферата.	Выполнены все требования к выполнению реферата, тема раскрыта, даны правильные ответы на дополнительные вопросы.
7-8	«Хорошо»		Выполнены основные требования к реферату, тема раскрыта не полностью, на дополнительные вопросы при защите даны неполные ответы.
5-6	«Удовлетворительно»		Требования к реферату выполнены не полностью, тема раскрыта не полностью, много упущений, допущены ошибки при ответе на дополнительные вопросы.
0-4	«Неудовлетворительно»		Реферат не выполнен; обнаруживается существенное непонимание в том, как его выполнять.

Зачет принимается в письменной форме на последнем практическом занятии. В задании к зачету даются два теоретических вопроса и одна задача. Студенту предоставляется 90 минут для письменного изложения ответов на вопросы и решению задачи. За ответы на теоретические вопросы студент максимально может получить 14 баллов. За решение задачи – максимально 6 баллов. Проходной балл на зачете – 12 баллов.

Методика оценивания ответов на зачете

Баллы	Оценка	Показатели	Критерии
12-20	«Зачтено»	<p>1. Полнота изложения теоретического материала;</p> <p>2. Полнота и правильность решения практического задания;</p> <p>3. Правильность и/или аргументированность изложения (последовательность действий);</p> <p>4. Самостоятельность ответа т.д.</p>	<p>Дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где обучающийся продемонстрировал знание дисциплины в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину, самостоятельно и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса, решил предложенные практические задания без ошибок.</p> <p>Дан развернутый ответ на поставленный вопрос, где студент демонстрирует знания, приобретенные на лекционных и семинарских занятиях, а также полученные посредством изучения обязательных учебных материалов по курсу, дает аргументированные ответы, приводит примеры. Однако допускается неточность в ответе. Решил предложенные практические задания с небольшими неточностями.</p> <p>Дан ответ, свидетельствующий в основном о знании процессов изучаемой дисциплины, отличающийся недостаточной</p>

			<p>глубиной и полнотой раскрытия темы, знанием основных вопросов теории, слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры. Допускается несколько ошибок в содержании ответа и решении практических заданий.</p>
0-11	«Не зачтено»		<p>Дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы. Выводы поверхностны. Решение практических заданий не выполнено, т.е. студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.</p>

**Лист актуализации оценочных материалов по дисциплине
«Технологии анализа больших данных»**

Оценочные материалы пересмотрены,
обсуждены и одобрены на заседании кафедры

Протокол от « _____ » _____ 20__ г. № _____

Зав. кафедрой _____

Оценочные материалы пересмотрены,
обсуждены и одобрены на заседании кафедры

Протокол от « _____ » _____ 20__ г. № _____

Зав. кафедрой _____

Оценочные материалы пересмотрены,
обсуждены и одобрены на заседании кафедры

Протокол от « _____ » _____ 20__ г. № _____

Зав. кафедрой _____

Оценочные материалы пересмотрены,
обсуждены и одобрены на заседании кафедры

Протокол от « _____ » _____ 20__ г. № _____

Зав. кафедрой _____