



Федеральное агентство по рыболовству
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Калининградский государственный технический университет»
(ФГБОУ ВО «КГТУ»)

УТВЕРЖДАЮ
Начальник УРОПСИ

Фонд оценочных средств
(приложение к рабочей программе дисциплины)

БОЛЬШИЕ ДАННЫЕ

основной профессиональной образовательной программы магистратуры
по направлению подготовки

38.04.01 ЭКОНОМИКА

Профиль программы

«ЭКОНОМИКА ОРГАНИЗАЦИЙ»

ИНСТИТУТ
РАЗРАБОТЧИК

отраслевой экономики и управления
кафедра экономической теории и инструментальных методов

1. РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными индикаторами достижения компетенций

Код и наименование компетенции	Индикаторы достижения компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ПК-6: Способен готовить аналитические материалы для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне	ПК-6.3: Готовит аналитические материалы по кластерам больших данных для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне	Большие данные	<p><u>Знать:</u></p> <ul style="list-style-type: none"> – методы прикладной статистики и теории вероятностей <p><u>Уметь:</u></p> <ul style="list-style-type: none"> – обрабатывать большие данные с помощью разных технологий (Hadoop, Hive, Spark, Hue, HBase, Kafka, Spark Streaming) <p><u>Владеть:</u></p> <ul style="list-style-type: none"> – методами машинного обучения – SQL и NoSQL СУБД – BI-системами (Power BI), формирую отчёты анализа данных; программируют на Python и работаю с библиотеками для анализа данных (NumPy, Matplotlib, scikit-learn)

2. ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПОЭТАПНОГО ФОРМИРОВАНИЯ РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДИСЦИПЛИНЫ (ТЕКУЩИЙ КОНТРОЛЬ) И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

2.1 Для оценки результатов освоения дисциплины используются:

- оценочные средства текущего контроля успеваемости;
- оценочные средства для промежуточной аттестации по дисциплине.

2.2 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания;
- задания по темам практических занятий.

2.3 К оценочным средствам для промежуточной аттестации по дисциплине, проводимой в форме экзамена, относятся:

- задания по курсовой работе;
- экзаменационные вопросы.

3. ОЦЕНОЧНЫЕ СРЕДСТВА ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ

3.1 Тестовые задания предназначены для оценки в рамках текущего контроля успеваемости знаний, приобретенных студентами на лекционных и практических занятиях и для измерения соответствующих индикаторов достижения компетенции.

Содержание теста определяется в соответствии с содержанием дисциплины пропорционально учебному времени, отведенному на изучение разделов, перечисленных в рабочей программе модуля.

Тестирование обучающихся проводится после рассмотрения на лекциях всех тем дисциплины.

Типовые варианты тестовых заданий приведены в Приложении №1.

3.2 Шкала оценивания тестовых заданий основана на четырехбалльной системе, которая реализована в программном обеспечении.

Оценка «отлично» выставляется при правильном выполнении не менее 90% заданий.

Оценка «хорошо» выставляется при правильном выполнении не менее 80% заданий.

Оценка «удовлетворительно» выставляется при правильном выполнении не менее 60% заданий.

Оценка «неудовлетворительно» выставляется при правильном выполнении менее 60% заданий.

Результаты измерений индикатора считаются положительными при правильном выполнении не менее 60% заданий.

3.3 В приложении № 2 приведены темы практических занятий и вопросы рассматриваемые на них. Задания для подготовки к практическим занятиям и материал, необходимый для подготовки к ним, в том числе показатели, критерии и шкалы оценивания результатов, представлены в учебно-методическом пособии, размещенном в электронной среде.

3.4 Шкала оценивания результатов выполнения заданий основана на двухбалльной системе «зачтено – не зачтено».

Магистрант, представивший самостоятельно выполненные отчетные материалы в соответствии с требованиями, способный ответить на контрольные вопросы теоретического введения, а также на вопросы об этапах выполнения работы и их результатах, получает оценку «зачтено».

Результаты измерений индикатора считаются положительными при положительной оценке за выполнение задания.

4 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

4.1 Учебным планом предусмотрено выполнение одной курсовой работы.

Каждому магистранту выдается индивидуальное задание на курсовую работу, как правило, в соответствии с темой его магистерской диссертации. Тематика курсовой работы приведена в Приложении № 3.

Оценка курсовой работы осуществляется по результатам её проверки преподавателем и по результатам её защиты магистрантом в соответствии с критериями, приведенными в таблице 2.

Кроме того, при оценке курсовой работы учитывается выполнения требований учебно-методического пособия по курсовой работе и методических указаний по оформлению учебных текстовых работ, размещенных в электронной среде.

4.2 Промежуточная аттестация по дисциплине проводится в форме экзамена.

К экзамену допускаются студенты, положительно аттестованные по результатам текущего контроля и курсовой работе.

Типовые экзаменационные вопросы приведены в Приложении № 4.

Представленные экзаменационные вопросы для проведения экзамена компонуются в билеты по два вопроса, относящиеся к различным темам и индикаторам двух разделов дисциплины. На усмотрение экзаменатора экзамен может быть проведен в письменной, устной или комбинированной форме, а также в форме тестирования. При наличии сомнений в отношении знаний и умений студента экзаменатор может (имеет право) задать дополнительные вопросы, а также дать дополнительное задание.

4.4 Критерии и шкала оценивания промежуточной аттестации.

Шкала итоговой аттестации по дисциплине, то есть оценивания результатов освоения дисциплины на экзамене, основана на четырехбалльной системе. Экзаменационная оценка («отлично», «хорошо», «удовлетворительно» или «неудовлетворительно») является экспертной и зависит от уровня освоения студентом тем дисциплины (наличия и сущности ошибок, допущенных студентом при ответе на экзаменационные вопросы).

При выставлении оценки по результатам сдачи экзамена используются следующие показатели и критерии оценивания.

Таблица 2 – Система оценок и критерии выставления оценки

Система оценок	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
Критерий	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
1 Системность и полнота знаний в отношении изучаемых объектов	Обладает частичными и разрозненными знаниями, которые не может научно-корректно связывать между собой (только некоторые из которых может связывать между собой)	Обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Обладает полнотой знаний и системным взглядом на изучаемый объект
2 Работа с информацией	Не в состоянии находить необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	Может найти необходимую информацию в рамках поставленной задачи	Может найти, интерпретировать и систематизировать необходимую информацию в рамках поставленной задачи	Может найти, систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
3. Научное осмысление изучаемого явления, процесса, объекта	Не может делать научно корректных выводов из имеющихся у него сведений, в состоянии проанализировать только некоторые из имеющихся у него сведений	В состоянии осуществлять научно корректный анализ предоставленной информации	В состоянии осуществлять систематический и научно корректный анализ предоставленной информации, вовлекает в исследование новые релевантные задачи данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы поставленной задачи
4. Освоение стандартных алгоритмов решения профессиональных задач	В состоянии решать только фрагменты поставленной задачи в соответствии с заданным алгоритмом, не освоил	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом,	Не только владеет алгоритмом и понимает его основы, но и предлагает новые решения в рамках поставленной

Система оценок	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
Критерий	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	предложенный алгоритм, допускает ошибки		понимает основы предложенного алгоритма	задачи

При проведении экзамена в форме тестирования используется шкала в п. 3.2.

Компетенции в той части, в которой они должны быть сформированы в рамках изучения дисциплины, могут считаться сформированными в случае, если магистрант получил на экзамене положительную оценку.

5. СВЕДЕНИЯ О ФОНДЕ ОЦЕНОЧНЫХ СРЕДСТВ И ЕГО СОГЛАСОВАНИИ

Фонд оценочных средств для аттестации по дисциплине Большие данные представляет собой компонент основной профессиональной образовательной программы магистратуры по направлению подготовки 38.04.01 Экономика (профиль «Экономика организаций»).

Фонд оценочных средств рассмотрен и одобрен на заседании кафедры Экономической теории и инструментальных методов (протокол № 8 от 01.04.2022 г.).

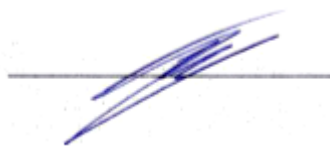
Заведующий кафедрой



Л. И. Сергеев

Фонд оценочных средств рассмотрен и одобрен на заседании кафедры экономики и финансов (протокол № 6 от 26.04.2022 г.)

Заведующий кафедрой



А.Г. Мнаçаканян

ТИПОВЫЕ ТЕСТОВЫЕ ЗАДАНИЯ

Вариант №1

1. Совокупность методов и средств повышения эффективности бизнеса за счет использования систем поддержки принятия решений, включающие программную и техническую инфраструктуру для сбора, хранения и аналитической обработки бизнес-данных носят общее название:

1. BI (Business Intelligence) =
2. Data Mining
3. OLAP (OnLine Analytical Processing)
4. Text mining

2. Алгоритм k-means (k-средних, k-ближайших соседей) является алгоритмом:

1. классификации
2. кластеризации =
3. регрессии
4. интерпретации

3. Задача регрессии является частным случаем задачи:

1. классификации =
2. кластеризации
3. ассоциации
4. интерпретации

4. Деревья решений — это группа алгоритмов:

1. классификации =
2. кластеризации
3. ассоциации
4. интерпретации

5. Построение (обучение) деревьев решений осуществляет алгоритм:

1. ID3 =
2. k-means
3. обратного распространения ошибок
4. рекурсии

6. Технология для хранения сложной структурированной и неструктурированной информации в компьютерных системах, обеспечивающая хранение фактов об окружающем мире, а механизм вывода которой использует различные формы логики для получения новых знаний называется:

1. базой знаний =
2. базой фактов
3. базой документов
4. базой данных

7. Уникальное значение или набор значений, которое однозначно идентифицирует запись в базе данных, называется:

1. первичным ключом =

2. внешним ключом
3. потенциальным ключом
4. составным ключом

8. Поле таблицы реляционной базы данных, содержащее значение, которое ссылается на первичный ключ другой таблицы и используется для связывания таблиц - это:

1. первичный ключ
2. внешний ключ =
3. естественный ключ
4. составным ключом

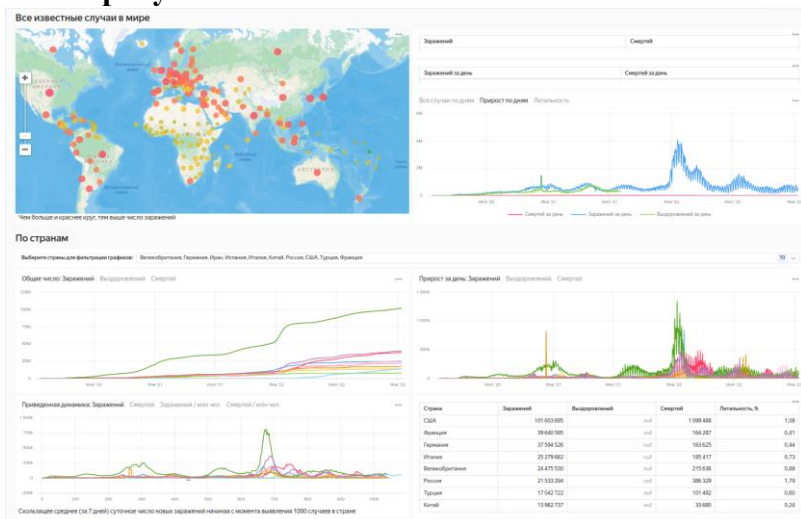
9. Средство визуализации данных, позволяющее объединить в одном визуальном информационном поле (интерфейсе) несколько информационных потоков в форме графиков, индикаторов, чартов, называют:

1. дашборд =
2. гистограмма
3. распределение
4. дендограмма

10. Для представления зависимости между тремя или более переменными может использоваться метод визуализации:

1. график рассеяния =
2. гистограмма
3. диаграмма размаха
4. в виде тепловых карт

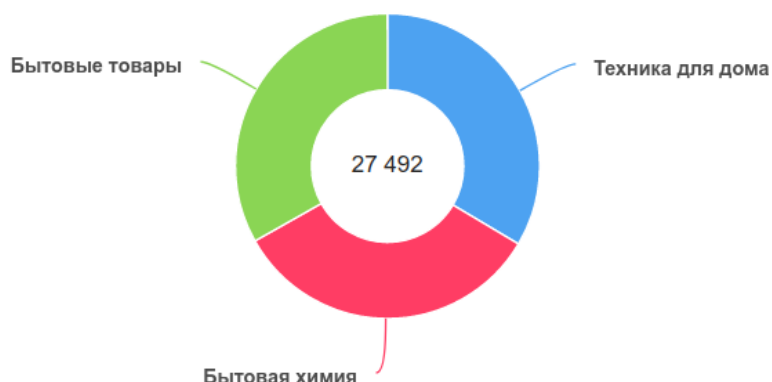
11. На рисунке



представлен пример

1. дашборда =
2. гистограммы
3. дендограммы
4. диаграммы Парето

12. На рисунке



представлен пример:

1. дашборда
2. линейчатой диаграммы
3. круговой диаграммы
4. кольцевой диаграммы =

13. Основная цель технологии MapReduce заключается в обеспечении

1. параллельной обработки больших наборов данных в вычислительном кластере =
2. хранения больших объемов структурированных и неструктурированных данных
3. выполнения сложного анализа данных
4. обработки данных в режиме реального времени

14. Дана таблица реляционной базы данных:

software_id	name	version	release_date	price	company	language	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	English	Operating System	Windows
2	Office 365	2019	2018-09-24	69.99	Microsoft	English	Productivity Suite	Windows, Mac, iOS, Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	English	Image Editing Software	Windows, Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	English	Operating System	Mac
5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	English	Web Browser	Windows, Mac, Linux,

								iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	English	Web Browser	Windows, Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	English	Operating System	Windows, Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	English	Music Streaming Service	Windows, Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	English	Version Control System	Windows, Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	English	Containerization Platform	Windows, Mac, Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQ L	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

При выполнении запроса

SELECT * FROM software WHERE price = '0'

будет получено ___ записей.

Введите число.

14

15. Дана таблица реляционной базы данных

software_id	name	version	release_date	price	company	language	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	English	Operating System	Windows

2	Office 365	2019	2018-09-24	69.99	Microsoft	English	Productivity Suite	Windows, Mac, iOS, Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	English	Image Editing Software	Windows, Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	English	Operating System	Mac
5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	English	Web Browser	Windows, Mac, Linux, iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	English	Web Browser	Windows, Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	English	Operating System	Windows, Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	English	Music Streaming Service	Windows, Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	English	Version Control System	Windows, Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	English	Containerization Platform	Windows, Mac, Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQL	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

При выполнении запроса

SELECT * FROM software WHERE release_date>'2021-01-01'

будет получено _____ записей.

Введите число.

10

Вариант №2

1. Технология получения информации из неструктурированных текстовых данных путём их преобразования в пригодный для дальнейшей работы набор структурированных данных, представленных в удобном для машинной обработки виде, называется:

1. BI (Business Intelligence)
2. Data Mining
3. OLAP (OnLine Analytical Processing)
4. Text mining =

2. Задача, заключающаяся в поиске оптимальных правил отнесения объекта к одному из классов, если число классов конечно, на основе имеющихся образцов (размеченного набора данных) называется:

1. классификация =
2. кластеризация
3. регрессия
4. интерпретация

3. Логистическая регрессия может использоваться для решения задачи:

1. бинарной классификации =
2. кластеризации
3. ассоциации
4. подобия

4. Алгоритм ID3 предназначен для:

1. построения (обучения) деревьев решений =
2. построения ассоциативных правил
3. обучения нейронных сетей
4. регрессии

5. Система для управления данными, которая используется для извлечения данных из различных источников, их преобразования и загрузки в целевую систему хранения данных, называется:

1. ETL (Extract, Transform, Load) системой =
2. системой Data Mining
3. системой поддержки принятия решений
4. BI-системой

6. Предметно-ориентированная база данных, содержащая данные по одному из направлений деятельности, хранение информации в которой оптимизировано с точки зрения решения конкретных задач, называется

1. витриной данных =
2. озером данных
3. базой знаний
4. ETL-системой

7. Таблица реляционной базы данных может иметь первичных ключей:

1. только один =
2. один или два

3. не более одного
4. любое количество

8. Комплекс методов представления исходной информации и результатов анализа данных в наиболее удобной для восприятия и интерпретации форме называют:

1. визуализаций =
2. кластеризацией
3. консолидацией
4. фрагментацией

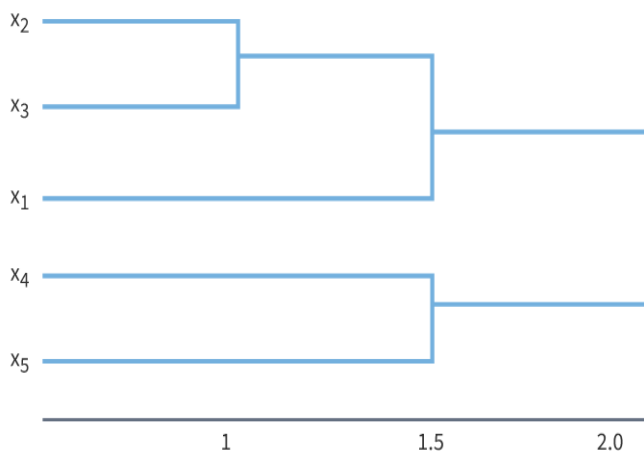
9. Визуально исследовать связь между двумя переменными позволяет метод визуализации:

1. график рассеяния =
2. гистограмма
3. диаграмма размаха
4. карта плотности

10. Spark – это инструмент для:

1. кластерных вычислений=
2. визуализации
3. работы с графиками
4. работы с базой знаний

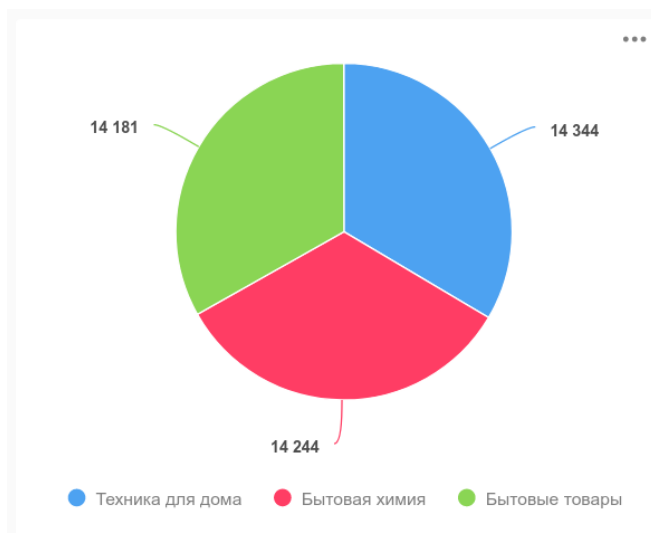
11. На рисунке



представлен пример

1. дашборда
2. гистограммы
3. дендограммы =
4. диаграммы Парето

12. На рисунке



представлен пример

1. дашборда
2. линейчатой диаграммы
3. круговой диаграммы =
4. кольцевой диаграммы

13. Функция Map в технологии MapReduce предназначена для:

1. группировки и агрегирования данных
2. сортировки данных
3. обработки и преобразования данных в промежуточную форму =
4. обработки данных в режиме реального времени

14. Дана таблица реляционной базы данных

software_id	name	version	release_date	price	company	language	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	English	Operating System	Windows
2	Office 365	2019	2018-09-24	69.99	Microsoft	English	Productivity Suite	Windows, Mac, iOS, Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	English	Image Editing Software	Windows, Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	English	Operating System	Mac
5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	English	Web Browser	Windows, Mac, Linux,

								iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	English	Web Browser	Windows, Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	English	Operating System	Windows, Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	English	Music Streaming Service	Windows, Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	English	Version Control System	Windows, Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	English	Containerization Platform	Windows, Mac, Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQL	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

При выполнении запроса:

SELECT * FROM software WHERE release_date BETWEEN '2021-01-01' AND '2021-01-31'

будет получено ____ записей.

Введите число.

1

15. Дана таблица реляционной базы данных:

software_id	name	version	release_date	price	company	language	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	English	Operating System	Windows
2	Office 365	2019	2018-09-24	69.99	Microsoft	English	Productivity Suite	Windows, Mac, iOS,

								Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	English	Image Editing Software	Windows, Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	English	Operating System	Mac
5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	English	Web Browser	Windows, Mac, Linux, iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	English	Web Browser	Windows, Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	English	Operating System	Windows, Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	English	Music Streaming Service	Windows, Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	English	Version Control System	Windows, Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	English	Containerization Platform	Windows, Mac, Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQL	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

При выполнении запроса:

SELECT * FROM software WHERE platform LIKE '%Mac%'

Будет получено _____ записей.

Введите число.

15

Вариант №3

Индикатор достижения компетенций ПКС-6.6: Готовит аналитические материалы по кластерам больших данных для оценки мероприятий в области экономической политики и принятия стратегических решений на микро- и макроуровне.

1. Направление информационных технологий, охватывающее всю область проблем, связанных с извлечением знаний из массивов данных называется:

1. Интеллектуальный анализ данных =
2. Базы данных и знаний
3. Статистический анализ данных
4. Машинное обучение

2. Задача, заключающаяся в поиске групп объектов в не размеченном исходном наборе данных, называется:

1. классификация
2. кластеризация =
3. регрессия
4. интерпретация

3. Если решается задача классификации на множестве двух классов, то классификация называется:

1. бинарной =
2. логистической
3. булевой
4. двумерной

4. Алгоритм CART предназначен для:

1. построения (обучения) деревьев решений =
2. построения ассоциативных правил
3. обучения нейронных сетей
4. интерполяции

5. Технология машинного обучения, когда нет ответов и требуется искать зависимости между объектами, называется:

1. самостоятельное обучение
2. обучение без учителя
3. обучение с учителем
4. обучение по зависимостям

6. База данных, в основе которой лежит модель данных в форме набора плоских таблиц, связь между которыми описывается в терминах первичного и внешнего ключа называется:

1. реляционной =
2. многомерной
3. объектно-ориентированной

4. предметно-ориентированной

7. Первичный ключ, включающий несколько полей таблицы реляционной базы данных, называется:

1. составным =
2. многозначным
3. индексом
4. производным

8. Дашборд – это:

1. Интерактивный инструмент для анализа данных =
2. Графический интерфейс для управления базой данных
3. Средство для сбора данных
4. Алгоритм для обработки данных

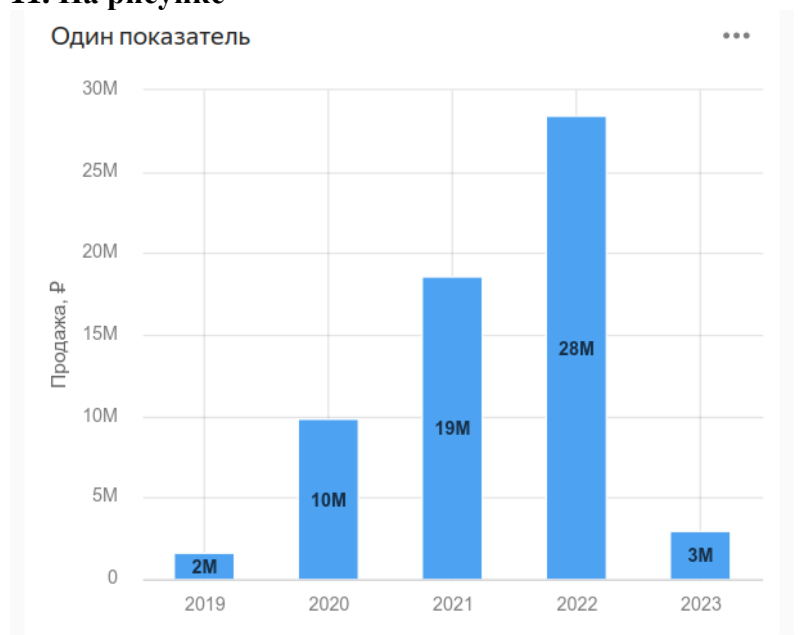
9. Для представления статистического распределения данных используется метод визуализации:

1. график рассеяния
2. гистограмма =
3. диаграмма размаха
4. карта плотности

10. Ассоциативное правило можно найти с помощью алгоритма:

1. apriori=
2. k-means
3. c-means
4. Hadoop

11. На рисунке

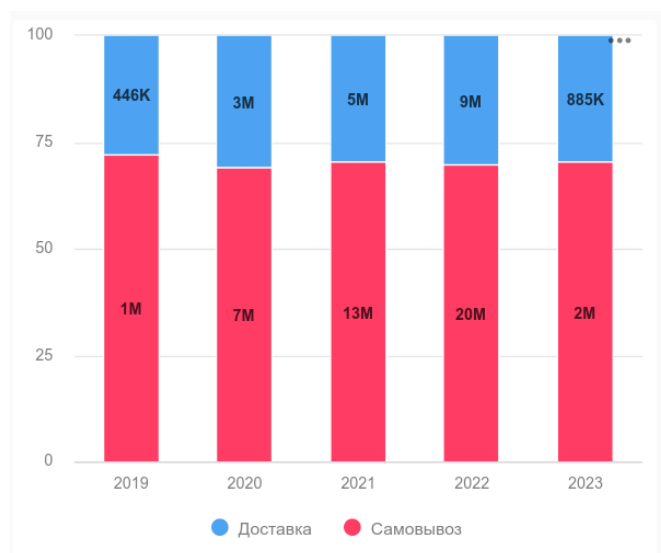


представлен пример:

1. дашборда

2. столбчатой диаграммы =
3. дендограммы
4. диаграммы Парето

12. На рисунке



представлен пример:

1. дашборда
2. линейчатой диаграммы
3. нормированной столбчатой диаграммы =
4. диаграммы Парето

13. Функция Reduce в технологии MapReduce предназначена для:

1. обработки данных в режиме реального времени
2. обработки данных в режиме реального времени
3. сортировки данных
4. консолидации результатов на основе промежуточных данных блоков Map=

14. Дана таблица реляционной базы данных:

software_id	name	version	release_date	price	company	language	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	English	Operating System	Windows
2	Office 365	2019	2018-09-24	69.99	Microsoft	English	Productivity Suite	Windows, Mac, iOS, Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	English	Image Editing Software	Windows, Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	English	Operating System	Mac

5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	English	Web Browser	Windows, Mac, Linux, iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	English	Web Browser	Windows, Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	English	Operating System	Windows, Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	English	Music Streaming Service	Windows, Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	English	Version Control System	Windows, Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	English	Containerization Platform	Windows, Mac, Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQL	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

При выполнении запроса

SELECT * FROM software Where type= 'Operating System'

будет получено ___ записей.

Введите число.

3

15. Дана таблица реляционной базы данных

so ft w ar e_ id	name	version	release_d ate	price	company	langua ge	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	Englis h	Operating System	Windows
2	Office 365	2019	2018-09-24	69.99	Microsoft	Englis h	Productivity Suite	Windows , Mac, iOS, Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	Englis h	Image Editing Software	Windows , Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	Englis h	Operating System	Mac
5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	Englis h	Source Code Editor	Windows , Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	Englis h	Source Code Editor	Windows , Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	Englis h	Web Browser	Windows , Mac, Linux, iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	Englis h	Web Browser	Windows , Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	Englis h	Operating System	Windows , Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	Englis h	Music Streaming Service	Windows , Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	Englis h	Version Control System	Windows , Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	Englis h	Containerization Platform	Windows , Mac,

								Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQL	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

При выполнении запроса

SELECT * FROM software WHERE type<> 'Operating System' AND company='Microsoft'

будет получено ___ записей.

Введите число.

4

Приложение № 2

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Практическое занятие № 1. ETL-система Loginom.

Цель: Изучение принципов консолидации и трансформации больших данных с использованием системы Loginom.

Материалы, оборудование, программное обеспечение: Loginom и/или доступ к серверу Loginom, доступ к api.hh.ru.

Критерии положительной оценки:

Предоставлены самостоятельно выполненные отчетные материалы в соответствии с требованиями. Студент способен ответить на контрольные вопросы теоретического введения, а также на вопросы об этапах выполнения работы и их результатах.

Планируемое время выполнения: 10 ак.ч.

Аудиторное время выполнения (под руководством преподавателя): 6 ак.ч.

Время самостоятельной подготовки: 4 ак.ч.

Задания:

1. Выполните задание «Множественный импорт файлов», опираясь на материал демопримера.
2. Выполните задание «Парсинг JSON», опираясь на материал демопримера.
3. Выполните задание «Поиск вакансий на hh.ru по ключевому слову», опираясь на материал демопримера.

Контрольные вопросы:

1. Чем импорт данных отличается от консолидации данных?
2. Какие форматы импорта поддерживает Loginom?
3. Дайте характеристику форматам xml, json и csv.
4. Опишите взаимодействие аналитической платформы Loginom с внешними сервисами.
5. Что такое API?
6. В чем заключается процедура «парсинга»?
7. Что такое REST-запрос?

Практическое занятие № 2. Предобработка и оценка качества больших данных.

Цель: Изучение основных задач предобработки и оценки качества данных в Loginom.

Материалы, оборудование, программное обеспечение: Loginom и/или доступ к серверу Loginom.

Критерии положительной оценки:

Предоставлены самостоятельно выполненные отчетные материалы в соответствии с требованиями. Студент способен ответить на контрольные вопросы теоретического введения, а также на вопросы об этапах выполнения работы и их результатах.

Планируемое время выполнения: 10 ак.ч.

Аудиторное время выполнения (под руководством преподавателя): 6 ак.ч.

Время самостоятельной подготовки: 4 ак.ч.

Задания:

1. Выполните задание «Разбор и очистка дат рождения», опираясь на материал демопримера и индивидуальный набор данных.
2. Выполните задание «Обнаружение аномальных значений временного ряда», опираясь на материал демопримера.

Контрольные вопросы:

1. Какие методы используются для предобработки больших данных?
2. Как можно оценить качество данных?
3. Как можно обработать пропущенные значения в данных?
4. Какие методы используются для уменьшения размерности данных?
5. Какие методы используются для обработки категориальных переменных?
6. Какие методы используются для нормализации данных?

Практическое занятие № 3. Визуализация данных.

Цель: Изучение принципов визуализации данных с использованием сервиса datalens.yandex.ru.

Материалы, оборудование, программное обеспечение: доступ к datalens.yandex.ru.

Критерии положительной оценки:

Предоставлены самостоятельно выполненные отчетные материалы в соответствии с требованиями. Студент способен ответить на контрольные вопросы теоретического введения, а также на вопросы об этапах выполнения работы и их результатах.

Планируемое время выполнения: 6 ак.ч. (очная форма обучения), 8 ак.ч. (заочная форма обучения),

Аудиторное время выполнения (под руководством преподавателя): 2 ак.ч. (очная форма обучения), 4 ак.ч. (заочная форма обучения),

Время самостоятельной подготовки: 4 ак.ч.

Задания:

1. Выполните задание «Анализ открытых данных ДТП на дорогах России», опираясь на материал демопримера.

Контрольные вопросы:

1. Что такое BI-система? Приведите примеры не менее 3 BI-систем.
2. В чем преимущество интерактивных панелей (дашбордов) перед статической презентацией?
3. В чем отличие «подключения» от «датасета» на платформе Yandex.DataLens.
4. В чем преимущества использования облачных платформ визуализации? Приведите не менее 5 преимуществ.
5. Предложите схему организации взаимодействия пользователей (аналитика, системного администратора, клиента, специалиста отдела продаж) платформы Yandex.DataLens при подготовке и использовании информационных панелей для некоторого модельного предприятия. Рассмотрите реальные кейсы внедрения.

Приложение №3

ТЕМАТИКА КУРСОВОЙ РАБОТЫ

Тема курсовой работы выбирается в первую очередь исходя из интересов профессиональной деятельности магистрантов.

Общая тема формулируется следующим образом «Визуальный анализ большого набора данных с использованием платформы Yandex.DataLens». Каждый магистрантов формулирует тему исходя из выбранного набора данных для анализа.

Например,

- «Визуальный анализ продаж сети магазинов с использованием платформы Yandex.DataLens»,
- «Визуальный анализ открытых данных ДТП на дорогах РФ с использованием платформы Yandex.DataLens»,
- «Визуальный анализ продаж и локаций пиццерий с использованием платформы Yandex.DataLens»,
- «Визуальный анализ статистики подкастов Яндекс Музыка с использованием платформы Yandex.DataLens».

Приложение № 4

ТИПОВЫЕ ЭКЗАМЕНАЦИОННЫЕ ВОПРОСЫ

1. Определение Больших данных. Источники больших данных.
2. Управление на основе данных. Значение больших данных для бизнеса и государства.
3. Цикл управления большими данными. Задачи анализа больших данных.
4. Технологии распределенных вычислений.
5. Обзор инструментальных средств работы с большими данными.
6. Понятие базы данных. СУБД.
7. Модели данных.
8. Хранилища данных, витрины данных.
9. Основы языка SQL.
10. Понятие пространства признаков. Метрики.
11. Введение в сокращение размерности. Сокращение числа признаков.
12. Оценка качества данных. Технологии и методы оценки качества данных.
13. Очистка и предобработка.
14. Фильтрация данных.
15. Обработка дубликатов и противоречий.
16. Выявление аномальных значений.
17. Восстановление пропущенных значений.
18. Задачи интеллектуального анализа.
19. Классификация. Логистическая регрессия,
20. Классификация. Искусственные нейронные сети
21. Кластеризации (k-means),
22. Регрессия. Алгоритм линейной регрессии
23. Ассоциации (Apriori).
24. Распределенные вычисления. Архитектура систем на основе BigData.
25. Технология MapReduce.
26. Виртуализация и облачные вычисления.
27. Проблемы информационной безопасности распределенных систем
28. Сторителлинг. Таблицы. Диаграмма.
29. Инфографика. Информационные панели.