

Частное образовательное учреждение высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ УНИВЕРСИТЕТ ТЕХНОЛОГИЙ
УПРАВЛЕНИЯ И ЭКОНОМИКИ»

РАССМОТРЕНО И ОДОБРЕНО

УТВЕРЖДАЮ

На заседании кафедры
информационных технологий и
математики
Протокол № 9 от 25.05.2023 г.

Первый проректор
С.В. Авдашкевич
28.06.2023

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина:	Б1.О.31 Анализ данных
Направление подготовки:	38.03.05 Бизнес-информатика
Направленность (профиль):	Цифровые решения для бизнеса
Уровень высшего образования:	Бакалавриат
Форма обучения:	очная, заочная, очно-заочная
Разработчики:	Кандидат технических наук, доцент Иванов С. А.

Санкт-Петербург
2023

1. Цели и задачи дисциплины:*Цель освоения дисциплины:*

изучить технологии анализа данных: OLAP, KDD, Data Mining и подготовки данных; дать представление об автоматизированных моделях анализа данных, применить методы анализа данных на примере решения задач сегментации, классификации, прогнозирования.

Задачи дисциплины:

- изучение понятийно-категориального аппарата в области углубленного анализа данных;
- формирование представлений об общей методологии консолидации, подготовки и анализа данных;
- обеспечение освоения современных методов OLAP, KDD, Data Mining;
- формирование навыков и умений, необходимых для создания и развития корпоративных аналитических систем.

2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы высшего образования

Планируемые результаты освоения ОП ВО (код и содержание компетенций)	Планируемые результаты обучения по ОП ВО (индикаторы достижения компетенций)	Примечание
ОПК-1 Способен проводить моделирование, анализ и совершенствование бизнес-процессов и информационно-технологической инфраструктуры предприятия в интересах достижения его стратегических целей с использованием современных методов и программного инструментария;	ОПК-1.1 Знает современные методы и программный инструментарий моделирования, анализа и совершенствования бизнес-процессов и информационно-технологической инфраструктуры предприятия.	-
	ОПК-1.2 Умеет применять современные методы и программный инструментарий моделирования, анализа и совершенствования бизнес-процессов и информационно-технологической инфраструктуры предприятия в интересах достижения его стратегических целей.	
	ОПК-1.3 Владеет современными методами и программным инструментарием моделирования, анализа и совершенствования бизнес-процессов и информационно-технологической инфраструктуры предприятия.	
ОПК-6 Способен выполнять отдельные задачи в рамках коллективной научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий.	ОПК-6.1 Знает способы реализации научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий.	-
	ОПК-6.2 Умеет применять коллективную форму реализации научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий.	
	ОПК-6.3 Владеет навыками поиска, выработки и применения новых решений в области информационно-коммуникационных технологий в рамках коллективной формы научно-исследовательской, проектной и учебно-профессиональной деятельности.	

Планируемые результаты обучения по ОП ВО (индикаторы достижения компетенций)	Планируемые результаты обучения по дисциплине
ОПК-1.1. Знает современные методы и программный инструментарий моделирования, анализа и совершенствования бизнес-процессов и информационно-технологической инфраструктуры предприятия.	Может выбирать современные инструменты для анализа разнородных данных.
ОПК-1.2. Умеет применять современные методы и программный инструментарий моделирования, анализа и совершенствования бизнес-процессов и информационно-технологической инфраструктуры предприятия в интересах достижения его стратегических целей.	Применяет современные инструменты для моделирования бизнес-процессов на основе проведенного анализа информационно-технологической инфраструктуры предприятия.
ОПК-1.3. Владеет современными методами и программным инструментарием моделирования, анализа и совершенствования бизнес-процессов и информационно-технологической инфраструктуры предприятия.	Проводит реинжиниринг бизнес-процессов на основе разработанных моделей с использованием интеллектуального анализа данных.
ОПК-6.1. Знает способы реализации научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий.	Проводит исследовательскую работу, направленную на первичный сбор разнородных данных для последующего анализа с помощью современных информационных технологий.
ОПК-6.2. Умеет применять коллективную форму реализации научно-исследовательской, проектной и учебно-профессиональной деятельности для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий.	Умеет работать в проектных группах с диверсификацией обязанностей и полномочий, выстраивать схему работы над задачей в коммуникации с другими участниками проекта.
ОПК-6.3. Владеет навыками поиска, выработки и применения новых решений в области информационно-коммуникационных технологий в рамках коллективной формы научно-исследовательской, проектной и учебно-профессиональной деятельности.	Выполняет поиск и первичную обработку данных для их последующего анализа и решения поставленной задачи.

3. Содержание, объем дисциплины и формы проведения занятий

№ п/п	Наименование темы дисциплины	Компетенции	Оценочные средства текущего контроля		
			ЗНАТЬ	УМЕТЬ	ВЛАДЕТЬ
			ОПК-1.1 ОПК-6.1	ОПК-1.2 ОПК-6.2	ОПК-1.3 ОПК-6.3
1	Введение в аналитику данных	ОПК-6	Тестирование №1 (10)	Доклад, сообщение/Реферат №1 (10)	Деловая и (или) ролевая игра/Кейс-задача №1 (20)
2	Подготовка данных	ОПК-1	Деловая и (или) ролевая игра/Кейс-задача №1 (20)	Деловая и (или) ролевая игра/Кейс-задача №1 (20)	Деловая и (или) ролевая игра/Кейс-задача №2 (20)
3	Визуализация данных	ОПК-1	Деловая и (или) ролевая игра/Кейс-задача №4 (20)	Деловая и (или) ролевая игра/Кейс-задача №4 (20)	Деловая и (или) ролевая игра/Кейс-задача №4 (20)
Количество баллов (100 баллов):			100		

Содержание учебного материала, лабораторные работы и практические занятия, курсовая работа

Тема 1: Введение в аналитику данных

Знакомство с анализом данных, его историей, областями применения аналитики данных, ее технологиями, задачами и инструментами. Изучение понятий, связанных с аналитикой данных, таких, как Data Mining, Big Data, Knowledge Discovery и пр.

Содержание учебного материала, лабораторные работы и практические занятия, курсовая работа
<p>Практические занятия/самостоятельная работа: Области применения аналитики данных.</p> <p>Лабораторная работа: -</p>
<p>Тема 2: Подготовка данных Группировка и преобразование даты: ETL-процесс. Преобразование данных как часть ETL-операций. Основные методы преобразования данных. Группировка данных. Функции агрегации. Преимущества группировки. Операции с датой и временем. Обогащение данных: Общие сведения об обогащении данных. Слияние (Внутреннее соединение. Левое и правое соединение. Полное соединение. Разность). Соединение. Дополнение данных. Объединение. Преобразование упорядоченных данных. Скользящее окно. Транспонирование данных: Структура данных. Представление данных. Некорректная структура. Понятие транспонирования. Транспонирование на этапе ETL. Обратное транспонирование.</p> <p>Практические занятия/самостоятельная работа: Основные методы преобразования данных.</p> <p>Лабораторная работа: -</p>
<p>Тема 3: Визуализация данных Введение в визуализацию: Визуализация и бизнес-аналитика. Цели и задачи визуализации. Этапы визуализации. Методы визуализации. Визуализаторы общего назначения: простые: Визуализаторы общего назначения. Таблицы. Графики. Диаграммы. Круговые диаграммы. Гистограммы. Визуализаторы общего назначения: сложные: Сложные визуализаторы общего назначения. OLAP-анализ. Географические карты. Тепловые карты. Плоское дерево. Диаграмма связей. Облако тегов. Пузырьковая диаграмма. Диаграмма рассеяния</p> <p>Практические занятия/самостоятельная работа: Этапы и методы визуализации.</p> <p>Лабораторная работа: -</p>
<p>Курсовая работа: не предусмотрено учебным планом</p>

Очная форма обучения

Вид учебной работы	Всего часов	Семестр 5
Аудиторные занятия (АЗ):	54	54
Лекционные занятия (Лек)	18	18
Лабораторные занятия (Лаб)	0	0
Практические занятия (Пр)	36	36
Самостоятельная работа студента (СР)	57	57
Курсовая работа	0	0
Другие виды самостоятельной работы*	57	57
Контроль самостоятельной работы (КСР)	6	6
Контактная работа (КоР)	60	60
Форма промежуточной аттестации	0	Экзамен
Подготовка к экзамену и сдача экзамена (СР, КоР)	27	27
Общая трудоемкость дисциплины, часы/ЗЕТ	144/4	144/4

* Подготовка к аудиторным занятиям, подготовка к зачету (при наличии)

№	Наименование темы дисциплины	Семестр/ Курс	Количество учебных часов				СР	Практическая подготовка
			В том числе по видам аудиторных занятий					
			Лек	Пр	Лаб			
1	Введение в аналитику данных	5	6	12	0	14	12	
2	Подготовка данных	5	6	12	0	14	12	
3	Визуализация данных	5	6	12	0	29	12	
Итого:			18	36	0	57	36	

* Практическая подготовка при реализации дисциплин организована путем проведения практических занятий и (или) выполнения лабораторных и (или) курсовых работ и предусматривает выполнение работ, связанных с будущей профессиональной деятельностью.

Заочная форма обучения

Вид учебной работы	Всего часов	Семестр 6
Аудиторные занятия (АЗ):	10	10

Вид учебной работы	Всего часов	Семестр 6
Лекционные занятия (Лек)	4	4
Лабораторные занятия (Лаб)	0	0
Практические занятия (Пр)	6	6
Самостоятельная работа студента (СР)	119	119
Курсовая работа	0	0
Другие виды самостоятельной работы*	119	119
Контроль самостоятельной работы (КСР)	6	6
Контактная работа (КоР)	16	16
Форма промежуточной аттестации	0	Экзамен
Подготовка к экзамену/зачету и сдача экзамена/зачета (СР, КоР)	9	9
Общая трудоемкость дисциплины, часы/ЗЕТ	144/4	144/4

* Подготовка к аудиторным занятиям

№	Наименование темы дисциплины	Семестр/ Курс	Количество учебных часов				СР	Практическая подготовка
			В том числе по видам аудиторных занятий					
			Лек	Пр	Лаб			
1	Введение в аналитику данных	6	2	2	0	36	12	
2	Подготовка данных	6	2	2	0	36	12	
3	Визуализация данных	6	0	2	0	47	12	
Итого:			4	6	0	119	36	

* Практическая подготовка при реализации дисциплин организована путем проведения практических занятий и (или) выполнения лабораторных и (или) курсовых работ и (или) путем выделения часов из часов, отведенных на самостоятельную работу, и предусматривает выполнение работ, связанных с будущей профессиональной деятельностью.

Очно-заочная форма обучения

Вид учебной работы	Всего часов	Семестр 6
Аудиторные занятия (АЗ):	18	18
Лекционные занятия (Лек)	8	8
Лабораторные занятия (Лаб)	0	0
Практические занятия (Пр)	10	10
Самостоятельная работа студента (СР)	86	86
Курсовая работа	0	0
Другие виды самостоятельной работы*	86	86
Контроль самостоятельной работы (КСР)	4	4
Контактная работа (КоР)	22	22
Форма промежуточной аттестации	0	Экзамен
Подготовка к экзамену и сдача экзамена (СР, КоР)	36	36
Общая трудоемкость дисциплины, часы/ЗЕТ	144/4	144/4

* Подготовка к аудиторным занятиям, подготовка к зачету (при наличии)

№	Наименование темы дисциплины	Семестр/ Курс	Количество учебных часов				СР	Практическая подготовка
			В том числе по видам аудиторных занятий					
			Лек	Пр	Лаб			
1	Введение в аналитику данных	6	2	2	0	28	12	
2	Подготовка данных	6	2	4	0	28	12	
3	Визуализация данных	6	4	4	0	30	12	
Итого:			8	10	0	86	36	

* Практическая подготовка при реализации дисциплин организована путем проведения практических занятий и (или) выполнения лабораторных и (или) курсовых работ и (или) путем выделения часов из часов, отведенных на самостоятельную работу, и предусматривает выполнение работ, связанных с будущей профессиональной деятельностью.

4. Способ реализации дисциплины

Без использования онлайн-курса.

5. Учебно-методическое обеспечение дисциплины:

Основная литература:

2. ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ. Учебник и практикум / Миркин Б. Г. - Национальный исследовательский университет «Высшая школа экономики» (г. Москва), 2022 г. - 174 с. - ISBN 978-5-9916-5009-0 – Режим доступа: <https://urait.ru/book/vvedenie-v-analiz-dannyh-469306>

3. АНАЛИЗ ДАННЫХ. Учебник для вузов / Под ред. Мхитаряна В.С. - Национальный исследовательский университет «Высшая школа экономики» (г. Москва), 2023 г. - 490 с. - ISBN 978-5-534-00616-2 – Режим доступа: <https://urait.ru/book/analiz-dannyh-511020>

3. АНАЛИЗ И ПРОЕКТИРОВАНИЕ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ UML. Учебное пособие для вузов / Галиаскаров Э. Г., Воробьев А. С. - Ивановский государственный химико-технологический университет (г. Иваново), 2022 г. - 125 с. - ISBN 978-5-534-14903-6 – Режим доступа: <https://urait.ru/book/analiz-i-proektirovanie-sistem-s-ispolzovaniem-uml-497207>

Дополнительная литература:

1. Павлов Л. А., Первова Н. В. — Структуры и алгоритмы обработки данных: учебник для вузов - Издательство Лань, 2021 г. - 256 с. - ISBN 978-5-8114-7259-8 – Режим доступа: <https://e.lanbook.com/book/156929>

2. ТЕОРИЯ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ И СИСТЕМ 2-е изд., пер. и доп. Учебник и практикум для вузов / Волкова В. Н. - Санкт-Петербургский политехнический университет Петра Великого (г. Санкт-Петербург), 2022 г. - 432 с. - ISBN 978-5-534-05621-1 – Режим доступа: <https://urait.ru/book/teoriya-informacionnyh-processov-i-sistem-489220>

3. Гулаков В. К., Трубаков А. О., Трубаков Е. О. — Структуры и алгоритмы обработки многомерных данных - Издательство Лань, 2021 г. - 356 с. - ISBN 978-5-8114-7965-8 – Режим доступа: <https://e.lanbook.com/book/169812>

6. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения

1. Операционная система
2. Пакет прикладных офисных программ
3. Антивирусное программное обеспечение
4. LMS Moodle
5. Вебинарная платформа
6. loginom academic

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», информационных справочных систем и профессиональных баз данных, необходимых для освоения дисциплины

1. ibooks.ru : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <https://ibooks.ru>. - Текст: электронный

2. Электронно-библиотечная система СПбУТУиЭ : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <http://libume.ru>. - Текст: электронный

3. Юрайт : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <https://urait.ru>. - Текст: электронный

4. eLibrary.ru : научная электронная библиотека [Электронный ресурс] : профессиональная база данных. - Режим доступа: <http://elibrary.ru>. - Текст: электронный

5. Архив научных журналов НЭИКОН [Электронный ресурс] : профессиональная база данных. - Режим доступа: arch.neicon.ru. - Текст: электронный

6. КиберЛенинка : научная электронная библиотека [Электронный ресурс] : информационная справочная система. - Режим доступа: <http://cyberleninka.ru>. - Текст: электронный

7. Лань : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <https://e.lanbook.com>. - Текст: электронный

8. it-world.ru [Электронный ресурс] : информационная справочная система . - Режим доступа: <https://www.it-world.ru>. - Текст: электронный

9. Connect: IT-технологии : информационная справочная система. - Режим доступа: <https://www.connect-wit.ru/>. - Текст: электронный

10. Цифровая экономика [Электронный ресурс] : информационная справочная система . - Режим доступа: <https://data-economy.ru/2024>. - Текст: электронный

11. Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации: профессиональная база данных. - Режим доступа: <https://digital.gov.ru>. - Текст: электронный

12. Министерство экономического развития Российской Федерации: профессиональная база данных. - Режим доступа: <http://economy.gov.ru>. - Текст: электронный

13. Федеральная служба по надзору в сфере связи, информационных технологий и массовых коммуникаций: профессиональная база данных . - Режим доступа: <https://rkn.gov.ru>. - Текст: электронный

8. Материально-техническое обеспечение дисциплины

1. Учебная аудитория для проведения занятий лекционного типа, семинарского типа - практических занятий, для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оборудованная: рабочими местами для обучающихся, оснащенными специальной мебелью; рабочим местом преподавателя, оснащенным специальной мебелью, персональным компьютером с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде Университета; техническими средствами обучения - мультимедийным оборудованием (проектор, экран, колонки) и маркерной доской; лицензионным программным обеспечением

2. Учебная аудитория для проведения занятий семинарского типа - практических занятий – компьютерный класс, оборудованный рабочими местами для обучающихся, оснащенными специальной мебелью, персональными компьютерами с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде Университета; рабочим местом преподавателя, оснащенным специальной мебелью, персональным компьютером с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде Университета; техническими средствами обучения - мультимедийным оборудованием (проектор, экран, колонки) и маркерной доской; лицензионным программным обеспечением

3. При применении электронного обучения, дистанционных образовательных технологий используются: виртуальные аналоги учебных аудиторий - вебинарные комнаты на вебинарных платформах, рабочее место преподавателя, оснащенное персональным компьютером (планшет, мобильное устройство) с возможностью подключения к сети «Интернет», доступом к электронной информационно-образовательной среде Университета и к информационно-образовательному portalу Университета imeos.ru, веб-камерой, микрофоном и гарнитурой (в т.ч. интегрированными в устройствами), программным обеспечением; рабочее место обучающегося оснащено персональным компьютером (планшет, мобильное устройство) с возможностью подключения к сети «Интернет», доступом к электронной информационно-образовательной среде Университета и к информационно-образовательному portalу Университета imeos.ru, веб-камерой, микрофоном и гарнитурой (в т.ч. интегрированными в устройства). Авторизация на информационно-образовательном portalе Университета imeos.ru и начало работы осуществляются с использованием персональной учетной записи (логина и

параля). Лицензионное программное обеспечение

4. Помещение для самостоятельной работы, оборудованное специальной мебелью, персональными компьютерами с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде Университета, лицензионным программным обеспечением

9. Оценочные материалы по дисциплине

Описание оценочных средств (показатели и критерии оценивания, шкалы оценивания) представлено в приложении к основной профессиональной образовательной программе «Каталог оценочных средств текущего контроля и промежуточной аттестации».

Процедуры оценивания знаний, умений, навыков и (или) опыта деятельности приводятся в соответствующих методических материалах и локальных нормативных актах Университета.

Для оценивания учебных достижений студентов в Университете действует балльно-рейтинговая система.

Если оценка, соответствующая набранной в семестре сумме рейтинговых баллов, удовлетворяет студента, то она является итоговой оценкой по дисциплине при проведении промежуточной аттестации в форме экзамена/зачета с оценкой/зачета.

Условием сдачи экзамена/зачета с оценкой/зачета с целью повышения итоговой оценки по дисциплине является сдача студентом экзамена, за который он получает экзаменационные баллы без учета баллов, полученных за текущий контроль:

Шкала оценивания учебных достижений по дисциплине, завершающейся зачетом без оценки

Баллы по дисциплине	60 и менее		61-73		74-90		91-100
Итоговая оценка по дисциплине	Незачет		Зачет				
Баллы в международной шкале ECTS с буквенным обозначением уровня	50 и менее	51-60	61-67	68-73	74-83	84-90	91-100
	F	Fx	E	D	C	B	A
Уровень сформированности компетенций	Не сформированы		Пороговый		Высокий		Повышенный

Шкала оценивания учебных достижений по дисциплине, завершающейся экзаменом/зачетом с оценкой

Баллы по дисциплине	60 и менее		61-73		74-90		91-100
Итоговая оценка по дисциплине	Неудовлетворительно		Удовлетворительно		Хорошо		Отлично
Баллы в международной шкале ECTS с буквенным обозначением уровня	<50	51-60	61-67	68-73	74-83	84-90	91-100
	F	Fx	E	D	C	B	A
Уровень сформированности компетенций	Не сформированы		Пороговый		Высокий		Повышенный

9.1. Типовые контрольные задания для текущего контроля Тестирование №1

1. Подход, при котором для решения задач выбираются готовые математические модели с известными параметрами, характерен для: А) вычислительной математики Б) разведочного анализа В) аналитики данных
2. Укажите наиболее точное определение бизнес-аналитики, которое сегодня подходит к аналитике данных в целом? А) Инструменты и приложения для поиска, анализа, моделирования и доставки информации, необходимой для принятия решений. Б) Технология в области программного и аппаратного обеспечения, интегрирует,

- организует, Q управляет и анализирует данные, характеризующиеся четырьмя характеристиками: объемом, разнообразием, изменчивостью и скоростью. В) Междисциплинарное направление информационных технологий, включающие все аспекты работы с данными с целью извлечения из них полезных знаний.
3. Какой столбец набора данных можно с большой вероятностью считать информативным для решения задачи анализа? А) Столбец, относительно которого выдвинута гипотеза о его влиянии на результат Б) Столбец, содержащий только одно уникальное значение В) Столбец, содержащий различные значения признака Г) Столбец с полностью уникальными значениями
 4. Укажите особенности настольных пакетов и библиотек с алгоритмами Data Mining. А) Отсутствие развитых инструментов интеграции данных Б) Обработка больших объемов данных В) Большое разнообразие алгоритмов анализа Г) Тесная интеграция с промышленными источниками данных
 5. Какие особенности малокодовых (low-code) систем в аналитике данных? А) Поток обработки данных представляет собой графическую диаграмму Б) При реализации логики часто требуется писать код на языке программирования В) Высокий порог входа для пользователя Г) Ограниченная функциональность
 6. Отметьте неверное высказывание по тематике Big Data. А) Объем данных в сотни терабайт или петабайт не позволяет легко хранить и управлять ими с помощью традиционных реляционных баз данных Б) Большинство всех данных Big Data являются структурированными В) Хранение огромного количества данных, описывающих некоторые легко наблюдаемые Г) события, не всегда приводит к выгодному пониманию реальности Д) Big Data, в контексте данных, обычно хранятся и организуются в распределенных файловых системах
 7. Какие способы защиты от копирования применяются в коммерческих редакциях Logiном? А) Аппаратный и программный ключ Sentinel Б) Текстовый файл с лицензией В) Аппаратный и программный ключ Guardant
 8. Как быстро создать независимую копию узла в рамках одного сценария с такими же настройками и связями? А) Командой Клонировать узел Б) Командой Копировать узел В) Командой Создать производный компонент...
 9. При каких настройках узла Узел 1 в Модуль 1 им можно будет воспользоваться в Модуль 2? Пакет состоит из Модуль 1 и Модуль 2. Выберите все варианты, комбинация которых даст желаемый результат. А) Модификатор доступа у Узел 1 — открытый Б) Модификатор доступа у Модуль 1 — закрытый В) Модификатор доступа у Модуль 1 — внутренний Г) Модификатор доступа у Узел 1 — внутренний Д) Модификатор доступа у Узел 1 — закрытый Е) Модификатор доступа у Модуль 1 — открытый
 10. При импорте текстового файла у части столбцов конвертация типов данных происходит с потерями. Что может помочь исправить ситуацию? А) Изменить вид данных Б) Изменить десятичный разделитель В) Изменить метку Г) Уменьшить значение параметра Кол-во строк для анализа Д) Изменить тип данных
 11. При добавлении нового узла в сценарий автосинхронизация его портов: А) включена Б) не включена
 12. Какие метаданные столбцов не принимаются во внимание алгоритмом автоматического связывания? А) Тип данных Б) Метка столбца В) Вид данных Г) Имя столбца
 13. Какая группа переменных имеет наивысший приоритет в сценарии? А) Пользователя Б) Системы В) Пакета Г) Сессии

Доклад, сообщение/Реферат №1

1. Проведение ABC-анализа в среде low-code.
2. Проведение RFM-анализа в среде low-code.
3. Проведение XYZ-анализа в среде low-code.

4. Решение задачи автоматизации в среде low-code.
5. Проведение анализа ассортимента в среде low-code.
6. Проведение анализа клиентской базы в среде low-code.
7. Проведение анализа конкурентов в среде low-code.
8. Проведение анализа отклонений в среде low-code.
9. Проведение анализа рисков в среде low-code.
10. Проведение классификации в среде low-code.
11. Оценка качества данных в среде low-code.
12. Проведение классификации с учетом издержек в среде low-code.
13. Проведение кластеризации в среде low-code.
14. Проведение кластерного анализа в среде low-code.
15. Проведение когортного анализа в среде low-code.
16. Решение задачи оптимизации в среде low-code.
17. Очистка данных в среде low-code.
18. Прогнозирование данных в среде low-code.
19. Проведение факторного анализа в среде low-code.
20. Аппроксимация в среде low-code.
21. Дисперсионный анализ в среде low-code.
22. Интерполяция в среде low-code.
23. Равномерное квантование в среде low-code.

Деловая и (или) ролевая игра/Кейс-задача №1.

Создадим простой сценарий, формирующий список 10 самых прибыльных групп товаров.

Сценарий выполнит действия:

- Импорт из файла Sales.lgd информации о продажах;
- Выделение 10 групп товаров с наибольшими суммами продаж;
- Экспорт полученных результатов.

1. Загрузите аналитическую платформу Loginom, создайте новый проект и сохраните его под именем лаб_раб_1.lgp.
2. Для того чтобы использовать в сценарии какой-либо компонент, его необходимо перенести мышью из панели компонентов в область построения сценария. Выберите в разделе Импорт компонент Loginom Data файл и перенесите его в область построения. При этом создастся узел сценария, выполняющий действия импорта. При клике мышкой на узле отобразятся иконки возможных действий.
3. Вызовите Мастер настройки. При выборе файла для импорта лучше использовать относительный путь, это означает, что файл с данными должен находиться в той же папке, что и файл проекта. Это позволит не перенастраивать узлы импорта при изменении местоположения папки на жестком диске и переносе сценариев с одного компьютера на другой.
4. После настройки узла выполните его, используя меню возможных действий. Теперь в выходном порте узла присутствуют импортированные данные, которые можно увидеть, выбрав Быстрый просмотр в контекстном меню порта.
5. Настройте следующие визуализаторы к узлу импорта: Таблица, Статистика.
6. Изучите визуализатор Статистика.

Деловая и (или) ролевая игра/Кейс-задача №2.

Цель: изучить программную среду хранилища данных, ознакомиться с архитектурой научиться создавать, и наполнять информацию из хранилища данных.

Хранилище данных (ХД) - это специально организованная база данных, ориентированная на решение задач анализа данных и поддержки принятия решений, обеспечивающая максимально быстрый и удобный доступ к информации. Warehouse 6 соответствует модели ROLAP (схема «снежинка»). Хранилище данных Warehouse включает в себя потоки данных, поступающие из различных источников, и специальный семантический слой, содержащий так называемые метаданные (данные о данных). Семантический слой и сами данные хранятся в одной базе

данных. Все данные в хранилище Warehouse хранятся в структурах типа «снежинка», где в центре расположены таблицы фактов, а «лучами» являются измерения, причем каждое измерение может ссылаться на другое измерение. Именно эта схема чаще всего встречается в хранилищах данных. Необходимо спроектировать хранилище с учетом того, что:

В таблице groups.txt Код группы является измерением, а Наименование группы - его атрибутом. В таблице produces.txt Код товара является измерением, а Наименование товара - его атрибутом, а Код группы - ссылкой на одноименное измерение.

В таблице stores.txt Код отдела является измерением, а Наименование отдела - его атрибутом.

В таблице sales.txt Дата является измерением, Отдел, Код товара и Код группы как было сказано выше – измерения. Час покупки - измерение, Количество и Сумма- факты, т.е. таблица sales.txt является описанием процесса продаж в трех аптеках.

Деловая и (или) ролевая игра/Кейс-задача №3.

Цель: изучить возможность поиска ассоциативных правил используя аналитическую платформу.

В последнее время растет интерес к методам «обнаружения знаний в базах данных». Большие объемы современных баз данных вызывают спрос на новые алгоритмы распознавания и обработки данных. Одним из распространенных аналитических методов обработки данных является аффинитивный анализ (англ: affinityanalysis), название произошедшее от английского слова affinity – близость, сходство. Метод определяет взаимные связи между событиями, происходящие совместно. Одним из применения аффинитивного анализа является анализ рыночной корзины (англ: marketbasketanalysis), цель которого – обнаружить ассоциации между различными данными, т.е. найти правила для количественного описания взаимной связи между двумя или более данными. Такие правила называются ассоциативными правилами (англ.: associationrules) и применяются в data mining.

Необходимо приложить ассоциативные правила для решения следующих задач (по выданным наборам данных):

1. Обнаружение наборов товаров, которые в супермаркетах часто покупаются вместе или никогда не покупаются вместе.
2. Определение доли клиентов, положительно относящихся к нововведениям в их обслуживании.
3. Определение профиля посетителя веб-ресурса.
4. Определение доли случаев, в которых новое лекарство показывает опасный побочный эффект.

Деловая и (или) ролевая игра/Кейс-задача №4.

Цель: научиться использовать метод кластерной обработки данных в виде самоорганизующихся карт Кохонена.

Импортируйте исходные данные из файла CreditSample.txt. Процесс построения карты Кохонена состоит из 10 этапов. Запустите мастер обработки, в котором в разделе «Data Mining» выберете способ обработки данных «Карта Кохонена», нажмите «Далее». В окне настройки назначения столбцов необходимо обозначить столбцы «Код» и «№ паспорта» как «Неиспользуемые» (так как значения этих столбцов уникальны, а это не позволит их классифицировать по общим признакам). Определите поле «Давать кредит» как «Выходное».

Используя различные отображения карты Кохонена, постройте 3-4 правила выдачи кредитов.

Содержание отчета:

1. Цель работы.
2. Краткое описание хода работы
3. Вид карт Кохонена
4. Ответы на вопросы.
5. Листинг программы
6. Заключение

9.2. Примерный перечень тем курсовой работы

Не предусмотрено учебным планом

9.3. Типовые контрольные задания для промежуточной аттестации: экзамен

Примерный перечень теоретических вопросов к экзамену (Вопрос №1)

1. Какие значения множества данных могут рассматриваться как аномальные?
2. Каково ожидаемое влияние аномальных значений на результаты анализа?
3. Как применяется визуальный анализ для выявления аномалий в одномерных и двумерных множествах данных?
4. Всегда ли аномальные значения являются нежелательными в данных?
5. Какие методы корректировки аномальных значений вам известны?
6. Что такое ETL-системы?
7. Чем вызвана необходимость использования ETL?
8. Каковы основные этапы процесса ETL и решаемые им задачи?
9. В чем отличие трансформации данных от предобработки и очистки?
10. С какими проблемами связана необходимость трансформации данных?
11. Каковы цели трансформации данных в аналитическом приложении?
12. Почему, несмотря на то, что трансформация данных производится на этапе консолидации данных, её необходимо применять и в аналитическом приложении?
13. Каково назначение систем OLTP и СППР?
14. Отличия OLTP-систем и СППР?
15. Обеспечивают ли ВХД поддержку исторических данных?
16. Что понимается в данных под пропущенным значением?
17. Почему пропущенные значения в анализируемых данных необходимо восстанавливать?
18. Каково происхождение пропусков в данных?
19. Каким требованиям должны удовлетворять алгоритмы восстановления пропущенных значений?
20. В каком случае пропущенные значения можно восстановить вручную?
21. Какие данные с точки зрения восстановления пропущенных значений являются упорядоченными, а какие – неупорядоченными?
22. В чем различие подходов между восстановлением пропусков в упорядоченных и неупорядоченных данных?
23. В чем преимущества и недостатки методики подстановки констант вместо пропущенных значений?

Примерный перечень практических заданий к экзамену (Вопрос №2)

1. Создайте пакет в аналитической платформе и импортируйте файл Товары.txt. Исключите из набора записи, в которых для товара отсутствует артикул. Сколько строк осталось в файле?
2. Создайте пакет в аналитической платформе и импортируйте файл Продажи.txt. С помощью функций компонента Калькулятор рассчитайте, сколько месяцев прошло от даты последней продажи каждого товара до 01.04.2018. Отсортируйте набор данных по количеству месяцев по убыванию. Какой код имеет тридцать седьмой товар?
3. Создайте пакет в аналитической платформе и импортируйте файл Численность_населения.txt. Выполните равномерное квантование двумя разными способами. Какое количество групп получилось в каждом случае? Почему это количество отличается/совпадает?
4. Необходимо построить сценарий в аналитической платформе, который ежедневно формирует отчет, показывающий 10 лидеров продаж по сумме продаж по итогам последних 10 дней. Результат необходимо визуализировать в виде упорядоченной по

убыванию столбчатой диаграммы. Необходимо произвести визуализацию в двух видах и сформировать соответствующие отчеты, а именно, с отображением сумм продаж, с отображением долей продаж в общей сумме продаж этих товаров. Исходные данные находятся в файле «Продажи», характеристики товаров в файле «Товары».

5. В файле представлены некоторые исходные данные, а именно столбцы «Вход1», «Вход2», «Выход». В данных присутствуют дубликаты и противоречия. Необходимо построить сценария в аналитической платформе, который исключает дубликаты и противоречия. Данные представлены в файле «Дубликаты и противоречия».
6. Коммерческий директор хочет иметь информацию о последних тенденциях в изменении суммовых объемов продаж по товарным группам. Для этого предлагается вычислить относительное изменение объемов продаж за последние 10 дней по отношению к объемам продаж за предыдущие 10 дней. Если это изменение менее -0,3, то товарной группе присваиваем категорию «Провал», от -0,3 до -0,1 – «Падение», от -0,1 до 0 – «Уменьшение», от 0 до 0,1 – «Увеличение», от 0,1 до 0,3 – «Подъем», свыше 0,3 – «Взлет». Необходимо построить соответствующий сценарий в аналитической платформе Deductor. Необходимые данные находятся в файлах «Продажи», «Товары», «Товарные группы».

Раздел билета	Компетенции	Планируемые результаты обучения по дисциплине	Количество баллов
Вопрос №1 Теоретический вопрос (проверяет знания («знать»), сформированные дисциплиной)	ОПК-1 ОПК-6	Может выбирать современные инструменты для анализа разнородных данных. Проводит исследовательскую работу, направленную на первичный сбор разнородных данных для последующего анализа с помощью современных информационных технологий.	40
Вопрос №2 Практическое задание (проверяет умения («уметь»), проверяет практические навыки («владеть»), сформированные дисциплиной)	ОПК-1 ОПК-6	Применяет современные инструменты для моделирования бизнес-процессов на основе проведенного анализа информационно-технологической инфраструктуры предприятия. Проводит реинжиниринг бизнес-процессов на основе разработанных моделей с использованием интеллектуального анализа данных. Умеет работать в проектных группах с диверсификацией обязанностей и полномочий, выстраивать схему работы над задачей в коммуникации с другими участниками проекта. Выполняет поиск и первичную обработку данных для их последующего анализа и решения поставленной задачи.	60