

Федеральное государственное образовательное бюджетное учреждение
высшего образования
«Финансовый университет при Правительстве Российской Федерации»
(Финансовый университет)

ЦЕНТР КОМПЕТЕНЦИЙ «ЦИФРОВАЯ ЭКОНОМИКА»

Обсуждено и одобрено
на Ученом совете Факультета

Протокол № 30
«21» марта 2023 г.



УТВЕРЖДАЮ
Проректор по цифровизации

Г.А. Остапенко

«21» марта 2023 г.

**ДОПОЛНИТЕЛЬНАЯ
ПРОФЕССИОНАЛЬНАЯ
ПРОГРАММА**

повышения квалификации

**«Технический аналитик
(Technical Analyst Data Science)»**

МОСКВА 2023

**Программа повышения квалификации
«Технический аналитик (Technical Analyst Data Science)»**

ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

Цель реализации программы: подготовки слушателей для получения компетенции в области информационных технологий в части машинного обучения и анализа данных для приобретения новой квалификации – технический аналитик - Technical Analyst Data Science.

Актуальность программы

Создание информационных технологий нового поколения, которые обеспечивают экономически эффективное извлечение полезной информации из больших объемов, является приоритетным направлением развития ИТ-сферы. Анализ данных нужен для управления и принятия решений, а также для разработки на их основе новых продуктов и услуг. Это делает актуальным подготовку специалистов, участвующих в процессе изучения данных, относящихся к конкретной ситуации, и вырабатывающих рекомендации, которые, дадут желаемые результаты. Для достижения этой цели технический аналитик будет использовать процесс технического анализа, который по сути является структурированным методом выявления и оценки ключевых факторов и элементов, и определения наиболее эффективного использования имеющихся ресурсов.

Разработка дополнительной профессиональной программы повышения квалификации на основе следующих стандартов:

1. Федеральные государственные образовательные стандарты высшего образования:

Дополнительная профессиональная программа (программа повышения квалификации) ИТ-профиля «Технический аналитик (Technical Analyst Data Science)» (далее – Программа) разработана:

- в соответствии с нормами Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- с учетом требований приказа Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», с изменениями, внесенными приказом Минобрнауки России от 15 ноября 2013 г. № 1244 «О внесении изменений в Порядок организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденный приказом Министерства образования и науки Российской Федерации от 1 июля 2013 г. № 499»;
- с учетом приказа Министерства образования и науки РФ от 23 августа 2017 г. № 816 «Об утверждении Порядка применения организациями,

осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»; паспорта федерального проекта «Искусственный интеллект» национальной программы «Цифровая экономика Российской Федерации»;

- с учетом постановления Правительства Российской Федерации от 13 мая 2021 г. № 729 «О мерах по реализации программы стратегического лидерства «Приоритет-2030» (в редакции постановления Правительства Российской Федерации от 14 марта 2022 г. № 357 «О внесении изменений в постановление Правительства Российской Федерации от 13 мая 2021 г. № 729»);
- с учетом приказа Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации от 28 февраля 2022 г. № 143 «Об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» и признании утратившими силу некоторых приказов Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» (далее – приказ Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации № 143);
- с учетом федеральных государственных образовательных стандартов (далее вместе – ФГОС ВО) по направлению подготовки 38.03.05 Бизнес-информатика (приказ Министерства науки и высшего образования РФ от 29.07.2020 г.);
- на основе анализа требований рынка труда в сфере искусственного интеллекта и анализа данных.

2. Наименования профессиональных стандартов, используемых при разработке программы повышения квалификации:

«Программист», утвержденный приказом Министерства труда и социальной защиты РФ от 18 ноября 2013 года №679н;

«Системный аналитик», утвержденный приказом Министерства труда и социальной защиты РФ от 28 октября 2014 года №809н;

«Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты РФ от 6 июля 2020 года №405н;

«Бизнес-аналитик», утвержденный приказом Министерства труда и соцзащиты РФ от 25 сентября 2018 года № 592н.

По окончании обучения слушатели должны обладать следующими профессиональными компетенциями:

- ПК-1. Применяет стандарты и методики бизнес-моделирования.
- ПК-2. Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.

– ПК-3. Использует большие данные: знает, понимает и обладает навыками использования разновидностей анализа данных; приемами анализа данных; проверкой гипотез о зависимости данных; методами сбора данных; с графическими данными; с текстовыми данными; с аудио данными; с видео данными и т.д.

– ПК-4. Применяет языки программирования для решения профессиональных задач - Python и SQL.

– ПК-5. Применяет СУБД.

По данной программе приобретаются компетенции универсальной модели компетенций в сфере искусственного интеллекта, разработанной РЭУ им. Г.В. Плеханова в 2021 году в рамках результат Федерального проекта «Искусственный интеллект»:

– ПК-1.р. - Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства для решения задач искусственного интеллекта;

– ПК-4.р. - Способен разрабатывать и применять методы машинного обучения для решения задач;

– ПК-5.р. - Способен использовать инструментальные средства для решения задач машинного обучения;

– ПК-6.р. - Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов;

– ПК-7.п. - Способен выполнять анализ больших данных;

– ПК-7.р. - Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта;

– ПК-8.и. - Способен использовать одну или несколько сквозных цифровых субтехнологий искусственного интеллекта;

– ПК-9.р. - Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта.

Область профессиональной деятельности:

06 Связь и информационно-коммуникационные технологии (в сферах: анализа, моделирования и формирования интегрального представления стратегий и целей, бизнес-процессов и информационно-технологической инфраструктуры предприятий различной отраслевой принадлежности и различных форм собственности, а также учреждений государственного и муниципального управления; стратегического планирования и управления развитием информационных систем и информационно-коммуникационных технологий управления предприятием; аналитической поддержки процессов принятия решений для управления предприятием);

07 Административно-управленческая и офисная деятельность (в сфере анализа, регламентирования, проектирования, оптимизации, автоматизации,

внедрения и контроля процессов и административных регламентов организаций с применением информационных технологий).

Рекомендуемые требования к уровню подготовленности, которым должны обладать граждане при поступлении на программу:

- Опыт работы с Microsoft Excel, опыт моделирования в MS Excel
- Понимание многомерных моделей данных (сводные таблицы в Excel, Power Query)

Базовые знания математического анализа, теории вероятностей, математической статистики

- Понимание бизнеса и экономической модели компании, цифровой экономики

- Умение быстро ориентироваться в цифрах, навыки работы с большими массивами данных.

Планируемые результаты обучения по программе повышения квалификации:

По итогам освоения программы слушатели должны:

Знать:

- о статистическом анализе: о методе многовариантного тестирования, корреляционном анализе, регрессионном анализе;
- о статистических методах: параметрические, непараметрические, управляемые, неуправляемые, полуправляемые, кластеризация;
- об алгоритмах машинного обучения: обучение с учителем, обучение без учителя, полуправляемое обучение, обучение с подкреплением;
- о машинном обучении: классификация, кластеризация, обнаружение выбросов, фильтрация;
- о современных языках запросов;
- об источниках информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования;
- о типах больших данных: метаданные, полуструктурированные, структурированные, неструктурированные;
- о видах источников данных: созданных человеком, созданных машинами;
- о типах данных и базовых алгоритмических конструкциях в Python, средствах языка для работы с массивами и строками, ввода/вывода данных через консоль и текстовые файлы, базовых понятиях ООП и принципах их использования, способах реализации механизмов ООП в Python, принципах создания приложений с графическим интерфейсом, назначениях и возможностях библиотеки Pygame;

- об основных понятиях и базовых алгоритмах в области анализа данных и машинного обучения, назначении и возможности библиотек Pandas, SEABORN, PlotLy, NumPy и Matplotlib;
- о теоретических основах анализа данных, технологиях OLAP и Data Mining;
- о методике KDD (Knowledge Discovery in Databases), которая используется при разработке моделей и решении стандартных задач Data Mining;
- о правилах и методах проектирования моделей данных, внедрения аналитических отчетов, информационных панелей мониторинга в компоненты ИТ инфраструктуры организации;
- об алгоритмических и программных методах проведения и интеллектуального анализа данных для принятия взвешенных решений.

Уметь:

- проводить сравнительный анализ методов и инструментальных средств анализа больших данных;
- планировать аналитические работы с использованием технологий больших данных;
- проводить анализ больших данных;
- осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных;
- применять язык SQL для работы по построению запросов, сбора и подготовки данных и визуализации данных;
- собирать, классифицировать, систематизировать и обеспечивать хранение и актуализацию информации бизнес-анализа;
- использовать современные информационные технологии аналитики для информационно-аналитического сопровождения деятельности организации;
- использовать интегрированную среду разработки для написания и отладки программного кода на языке Python, описывать классы и иерархию связей между ними, реализовывать механизмы перегрузки операторов, документировать классы, устанавливать и подключать к проекту библиотеки, использовать структуры данных и методы библиотек Pygame, Pandas, SEABORN, PlotLy, NumPy и Matplotlib;
- решать задачи профессиональной деятельности, используя OLAP технологии аналитики;
- использовать современные информационные технологии аналитики для информационно-аналитического сопровождения деятельности;
- работать с многомерными моделями.

Владеть:

- навыками использования технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенной и параллельной обработки данных, вычислений в оперативной памяти;
- навыками использования технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, планирование экспериментов, карты контроля качества;
- анализом решений с точки зрения достижения целевых показателей решений;
- навыками работы с сырыми данными, использования SQL для интеграции и очистки данных;
- навыками применения технологий машинного обучения для решения анализа данных и принятия решений;
- навыками работы в интегрированной среде разработки для Python, разработки программ с использованием механизмов ООП, навыками разработки графического интерфейса пользователя, навыками решения простейших задач классификации и анализа данных на языке Python;
- навыками использования инструментов и технологий выявления, сбора и анализа информации бизнес-анализа для формирования возможных решений;
- навыками использования инструментов по формированию аналитических отчетов, дашбордов и «повествований» для поддержки принятия решений.

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Центр компетенций «Цифровая экономика»

Обсуждено и одобрено
на Ученом совете Факультета

Протокол № 30
«21» марта 2023 г.



УТВЕРЖДАЮ
Проректор по цифровизации

Г.А. Остапенко

«21» марта 2023 г.

УЧЕБНЫЙ ПЛАН

программы повышения квалификации

**«Технический аналитик
(Technical Analyst Data Science)»**

Требования к уровню образования слушателей	Имеющие высшее или среднее профессиональное образование и (или) получающие высшее или среднее профессиональное образование
Категория слушателей	Предприниматели, сотрудники и собственники ИТ компаний, имеющие высшее образование или среднее профессиональное (либо получающие высшее или среднее профессиональное образование), а также высоко мотивированные специалисты из других сфер и студенты, обучающиеся сфере информационных технологий, а также по иным специальностям, которые заинтересованы в приобретении новой квалификации «Технический аналитик (Technical Analyst Data Science)»
Срок обучения	256 часов
Форма обучения	Очная – заочная с применением электронного обучения и дистанционных образовательных технологий
Режим занятий	4-6 часов в день

№ п/п	Наименование дисциплины, модуля, темы	Трудоемкость		В том числе				Форма контроля
		В зачетных единицах	В часах	Всего	Контактная работа ¹		Самостоятельная работа*	
					Лекции	Практические		
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
1.	МОДУЛЬ 1 Введение в аналитику данных Направление - «Технический аналитик (Technical Analyst Data Science)»		36	22	10	12	12	2, тест
2.	Промежуточная аттестация		2	-	-	-		Тест
3.	Итого по модулю 1		38	22	10	12		
4.	МОДУЛЬ 2 Технологии хранения и обработки данных		60	39	20	19	19	2, тест
5.	Промежуточная аттестация		2	-	-	-		Тест
6.	Итого по модулю 2		62	39	20	19		
7.	МОДУЛЬ 3 Языковые средства для аналитики данных		52	32	14	18	20	2, тест
8.	Промежуточная аттестация		2	-	-	-		Тест
9.	Итого по модулю 3		54	32	14	18		
10.	МОДУЛЬ 4 Использование языка Python для анализа данных		34	22	8	14	12	2, тест
11.	Промежуточная аттестация		2	-	-	-		Тест
12.	Итого по модулю 4		36	22	8	14		
13.	МОДУЛЬ 5 Технический анализ данных и подготовка данных к визуализации и анализу		58	38	10	28	20	2, тест
14.	Промежуточная аттестация		2	-	-	-		Тест
15.	Итого по модулю 5		60	38	10	28		
16.	Итоговая аттестация (Экзамен в форме тестирования и решение практического кейса)		10	-	-	-		10, Зачет
17.	Итого		256	153	62	91	83	20
18.	Общая трудоемкость программы		256	153	62	91	83	20

Разработчики программы:

Сурова Надежда Юрьевна – директор Центра Компетенций «Цифровая экономика» Финансового университета, кандидат экономических наук, доцент, ведущий архитектор программного обеспечения Digital Auto Systems, Директор Международного евразийского индустриального фонда, Член Совета Совета

¹ С применением дистанционных образовательных технологий и электронного обучения

Федерации ФС РФ по цифровой экономике, Член экспертного совета по импортозамещению ИТ Государственной Думы РФ.

Мещеряков Роман Валерьевич – Доктор технических наук, Профессор Российской академии наук, Главный научный сотрудник, заведующий лабораторией киберфизических систем, директор центра Интеллектуальных робототехнических систем.

Петросов Давид Арегович, Финансовый университет, руководитель Департамента анализа данных, принятия решений и финансовых технологий, кандидат технических наук, доцент

Захарова (Гельгорн) Алёна Александровна – Доктор технических наук - специальность 05.13.01 «Системный анализ, управление и обработка информации», профессор, Эксперт Федерального реестра экспертов по информационным технологиям.

В реализации программы принимают участие специалисты и преподаватели Финансового университета, РАН, других вузов, а также приглашенные бизнес – практики и ведущие специалисты в профильной сфере из компаний партнеров программы: Международного евразийского индустриального фонда, Инжинирингового центра "Гипертех", Центра развития базисных инноваций и прототипирования, ООО «Актив Бизнес консалтинг», АО «ВКОНТАКТЕ», ПАО «МТС», ООО «Открытые системы», ООО «ЭР – Телеком», ООО «Вымпелком» (Билайн), ООО «Цифровая Формула Доверия» (ООО «Цифровая Формула Доверия») другие.

Директор Центра Компетенций
«Цифровая экономика»



Н.Ю. Сурова

«21» марта 2023 г.

**Учебно-тематический план программы
повышения квалификации
«Технический аналитик
(Technical Analyst Data Science)»**

№ п/п	Наименование дисциплины, модуля, темы	Трудоем- кость		В том числе				Форма контроля
		В зачетных единицах	В часах	Всего	Контактная работа ²		Самостоятельная работа*	
					Лекции	Практичес- кие		
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
1.	МОДУЛЬ 1 Введение в аналитику данных Направление - «Технический аналитик (Technical Analyst Data Science)»		36	22	10	12	12	2, тест
2.	Тема 1. Профессия «Технический аналитик: область деятельности, необходимые хард и софт скилы		6	4	2	2	2	-
3.	Тема 2. Основные понятия и методы моделирования бизнес процессов		6	4	2	2	2	-
4.	Тема 3. Бизнес-анализ: предиктивная аналитика, предписывающий анализ		6	4	2	2	2	-
5.	Тема 4. Теория вероятности и математическая статистика в аналитике данных		6	4	2	2	2	-
6.	Тема 5. Обзор языковых, программных, инструментальных средств аналитики данных		6	4	2	2	2	-
7.	Тема 6. Обзор методик обнаружения новых знаний, работа с большими данными		4	2	-	2	2	-
8.	Промежуточная аттестация		2	-	-	-		Тест
9.	Итого по модулю 1		36	22	10	12		
10.	МОДУЛЬ 2 Технологии хранения и обработки данных		60	39	20	19	19	2, тест
11.	Тема 1. Современные технологии хранения и обработки данных		18	12	6	6	6	
12.	Тема 2. Статистические методы анализа и обработки данных		18	12	6	6	6	
13.	Тема 3. Обработка данных с использованием табличных процессоров. Нейронные сети		22	15	8	7	7	

² С применением дистанционных образовательных технологий и электронного обучения

14.	Промежуточная аттестация		2	-	-	-		Тест
15.	Итого по модулю 2		60	39	20	19		
16.	МОДУЛЬ 3 Языковые средства для аналитики данных		52	32	14	18	20	2, тест
17.	Тема 1. Введение в языки запросов		20	10	4	6	10	
18.	Тема 2. Построение запросов SQL		15	10	4	6	5	
19.	Тема 3. SQL в работе с большими данными		17	12	6	6	5	
20.	Промежуточная аттестация		2	-	-	-		Тест
21.	Итого по модулю 3		52	32	14	18		
22.	МОДУЛЬ 4 Использование языка Python для анализа данных		34	22	8	14	12	2, тест
23.	Тема 1. Основные конструкции языка Python		14	10	4	6	4	
24.	Тема 2. Взаимодействие языка Python и SQL		10	6	2	4	4	
25.	Тема 3. Возможности Python для анализа данных		10	6	2	4	4	
26.	Промежуточная аттестация		2	-	-	-		Тест
27.	Итого по модулю 4		34	22	8	14		
28.	МОДУЛЬ 5 Технический анализ данных и подготовка данных к визуализации и анализу		58	38	10	28	20	2, тест
29.	Тема 1. Технический анализ данных и подготовка данных к визуализации и анализу. Обзор систем визуального анализа данных.		28	18	4	14	10	
30.	Тема 2. OLAP-анализ		30	20	6	14	10	
31.	Промежуточная аттестация		2	-	-	-		Тест
32.	Итого по модулю 5		58	38	10	28		
33.	Итоговая аттестация (Зачет в форме тестирования и решение практического кейса)		10	-	-	-		Зачет
34.	Итого		256	153	62	91	83	20
35.	Общая трудоемкость программы		256	153	62	91	83	20

Федеральное государственное образовательное бюджетное учреждение
высшего образования
"Финансовый университет при Правительстве Российской Федерации"
(Финансовый университет)
Центр компетенций «Цифровая экономика»

Календарный учебный график

программы повышения квалификации
«Технический аналитик (Technical Analyst Data Science)»

Объем программы – 256 часа.

Продолжительность обучения – 3,5 месяца

Форма обучения – Очно – заочная с применением электронного обучения и дистанционных образовательных технологий.

Образовательный процесс по программе может осуществляться в течение всего учебного года. Занятия проводятся по мере комплектования учебных групп.

№ п\п	Наименование дисциплин (модулей), тем	Месяц			КР	Л	ПЗ	СР	ПА	ИА	Всего
		1	2	3							
1	Модуль 1 Введение в аналитику данных Направление - «Технический аналитик (Technical Analyst Data Science)»	12	12	14	22	10	12	12	2		38
2	МОДУЛЬ 2 Технологии хранения и обработки данных	20	22	20	39	20	19	19	2		62
3	МОДУЛЬ 3 Языковые средства для аналитики данных	18	18	18	32	14	18	20	2		54
4	МОДУЛЬ 4 Использование языка Python для анализа данных	13	13	13	22	8	14	12	2		36

5	МОДУЛЬ 5 Технический анализ данных и подготовка данных к визуализации и анализу	20	20	20	38	10	28	20	2		60
6	Итоговая аттестация (зачет в форме тестирования и решения практического кейса)			10						10	
7	Итого	83	85	95	153	62	91	83	10	10	256
8	Общая трудоемкость программы	83	85	95	153	62	91	83	10	10	256

Условные обозначения	
ПА	Промежуточная аттестация
Л	Лекция
ПЗ	Практические занятия
ИА	Итоговая аттестация
КР	Контактная работа

Директор Центра Компетенций
«Цифровая экономика»

_____ Н.Ю. Сурова

«21» марта 2023 г.

**Рабочие программы модулей учебного курса программы повышения
квалификации
«Технический аналитик (Technical Analyst Data Science)»**

Программа повышения квалификации состоит из 5 учебных модулей:

Модуль 1 Введение в аналитику данных Направление - «Технический аналитик (Technical Analyst Data Science)»
Модуль 2 Технологии хранения и обработки данных
Модуль 3 Языковые средства для аналитики данных
Модуль 4 Использование языка Python для анализа данных
Модуль 5 Технический анализ данных и подготовка данных к визуализации и анализу

Рабочая программа модуля 1

«Введение в аналитику данных»

Направление - «Технический аналитик (Technical Analyst Data Science)»

Цель освоения модуля 1 – предоставление знаний по новой профессии: область деятельности, необходимые хард и софт скилы, обзора языковых, программных, инструментальных средств аналитики данных и методик обнаружения новых знаний, работа с большими данными, предоставление понимания методов моделирования бизнес- процессов, теории вероятности и математическая статистика в аналитике данных, приобретение слушателями компетенций по бизнес-анализу: предиктивной аналитике, предписывающему анализу.

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 1:

- ПК-1. Применяет стандарты и методики бизнес-моделирования.
- ПК-2. Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.

Планируемые результаты обучения по модулю 1

По итогам освоения модуля слушатели должны:

Знать:

- область деятельности и необходимые хард и софт скилы профессии;
- основные понятия и методы моделирования бизнес- процессов;
- о статистическом анализе: о методе многовариантного тестирования, корреляционном анализе, регрессионном анализе;

- о статистических методах: параметрические, непараметрические, управляемые, неуправляемые, полууправляемые, кластеризация;
- об алгоритмах машинного обучения: обучение с учителем, обучение без учителя, полууправляемое обучение, обучение с подкреплением;
- о машинном обучении: классификация, кластеризация, обнаружение выбросов, фильтрация;
- о современных языках запросов;
- об источниках информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования;

Уметь:

- проводить сравнительный анализ методов и инструментальных средств анализа больших данных;
- планировать аналитические работы с использованием технологий больших данных;

Владеть:

- навыками использования инструментов бизнес-анализа: предиктивной аналитике, предписывающему анализу.

Учебно-тематический план модуля 1

«Введение в аналитику данных»

Направление - «Технический аналитик (Technical Analyst Data Science)»

№ п/п	Наименование дисциплины, модуля, темы	Трудоемкость		В том числе				Форма контроля
		В зачетных единицах	В часах	Всего	из них		Самостоятельная работа*	
					Лекции	Практические		
1	2	3	4	5	6	7	8	9
1.	Тема 1. Введение в аналитику данных. Направление - «Технический аналитик (Technical Analyst Data Science)»		6	4	2	2	2	-

³ С применением дистанционных образовательных технологий и электронного обучения

2.	Тема 2. Основные понятия и методы моделирования бизнес процессов		6	4	2	2	2	-
3.	Тема 3. Бизнес-анализ: предиктивная аналитика, предписывающий анализ		6	4	2	2	2	-
4.	Тема 4. Теория вероятности и математическая статистика в аналитике данных		6	4	2	2	2	-
5.	Тема 5. Обзор языковых, программных, инструментальных средств аналитики данных		6	4	2	2	2	-
6.	Тема 6. Обзор методик обнаружения новых знаний, работа с большими данными		4	2	-	2	2	-
7.	Промежуточная аттестация		2	2	-	2		-
8.	Итого по модулю 1		36	22	10	12	12	2, тест

Тема 1. Профессия «Технический аналитик (Technical Analyst Data Science)»: область деятельности, необходимые хард и софт скилы

Профессия: область деятельности, необходимые хард и софт скилы. Управление пользователями и ролями пользователей. Проектирование и поддержка системы данных и базы данных. Наборы данных о вводе в эксплуатацию и выводе из эксплуатации. Техническая экспертиза в области структуры хранения данных, интеллектуального анализа и очистки данных. Поддержка инициатив по обеспечению целостности и нормализации данных. Сбор и анализ требований заказчиков к отчетности. Получение данных с помощью языка запросов SQL. Применение в работе ключевых математических методов и основ статистики. Очистка и трансформация данных с помощью Python. Прогнозирование событий на основе данных. Анализ результатов кампаний, исследований и тестирования продуктовых гипотез. Способность создавать аналитические решения и представлять их бизнесу. Статистические инструменты для интерпретации наборов данных, обращая внимание на определенные тенденции и закономерности. Примеры аналитических навыков для каждого аналитика данных.

Тема 2. Основные понятия и методы моделирования бизнес процессов

Моделирование бизнес-процессов: основные понятия Business Process Modeling. Действия, данные, события, материалы. Системное моделирование бизнес-процессов в виде блок-схем, диаграмм, таблиц, сценариев и т.д. Способы, выбранные для наглядного отображения элементов. К чему приводит отсутствие формализованных бизнес-процессов. Нормирование бизнес-процессов. Гибкость процессов. Анализ бизнес-процессов для формирования культуры инноваций и изменений. Возможность настраивать бизнес-операции позволяет компании развиваться в условиях

технологических изменений; прозрачность. Виды и принципы моделирования бизнес-процессов. Функциональное моделирование бизнес-процессов описывает их в виде функций, которые четко структурированы и взаимосвязаны между собой. Объектное моделирование: бизнес-процессы отображаются как набор объектов, взаимодействующих между собой. Имитационное моделирование — в этом случае процессы представляют через примеры их поведения в разных условиях, анализируя свойства в динамике. Ориентация на эталонные и референтные модели как базу для описания бизнес-процессов. Моделирование «сверху вниз». Соизмеримость процессов по сложности (составу) и по значимости. Целостность описания процесса: задание его названия, последовательности функций, участников процесса, используемых ресурсов. Стадии моделирования бизнес-процессов. Тестирование построенной модели — внедрение ее в деятельность компании, оценка результатов, внесение изменений. Улучшение построенной модели — в процессе использования модель необходимо продолжать анализировать и совершенствовать. Методология и инструментарий моделирования бизнес-процессов IDEF — класс методов (IDEF0, IDEF1 и т.д.), основанных на методологии SADT. VAD — нотация дает общий взгляд на бизнес-процессы, которые непосредственно участвуют в создании ценности, т.е. продукта или услуги. EPC — нотация позволяет создать диаграмму процессов нижнего уровня, в которой для всех событий и функций определены участники, материальные и информационные потоки, стартовые и финишные точки. BPMN — нотация, которая моделирует шаги запланированного бизнес-процесса от начала до завершения. Data Flow Diagram отображает передачу информации (не материалов) между операциями в рамках процесса. Role Activity Diagram используют для моделирования процесса как совокупности ролей, имеющих определенные функции, и их взаимодействия. Flow Chart Diagram строится с помощью набора символов, которые обозначают элементы процесса: процедуры, инструменты, данные и т.д. Сети Петри - граф, вершины которого — это действия процесса, а дуги — события, определяющие изменение состояния процесса. Ряд программных продуктов, которые могут быть использованы в качестве инструментов моделирования бизнес-процессов с применением описанных методов: ARIS, Business Studio, MS Visio, Bizagi Process Modeler и др. Моделирование бизнес-процессов: проект из практики. Результаты бизнес-моделирования.

Тема 3. Бизнес-анализ: предиктивная аналитика, предписывающий Анализ

Рассматривают основные понятие бизнес-анализа, анализ, как метод научного познания, дается определение бизнес-анализа как деятельности, согласно своду знаний о бизнес-анализе (Business Analysis Body of Knowledge 2.0) (BABOK), содержание бизнес-аналитики как концепции изучения и понимания компанией своего бизнеса, структура, функции, составные части бизнес-анализа, приводятся методика накопления и систематизации бизнес-

информации, доступ к информации и вопросы этики. Предписывающий анализ.

Рассматривается практический кейс «Аналитика продажи товаров торговой сети»

Описание:

На основе дата-сета Transactions_diy.lgd необходимо выполнить следующие практические задания:

- Расчет ТОП-10 товаров по количеству продаж в торговой сети
- Определение количества уникальных товаров в дата-сете с транзакциями
- Определение товара, лидирующего по продажам за последние 6 месяцев
- Количество уникальных чеков в августе 2018 года
- Построение ABC-XYZ матрицы по данным за последние 6 месяцев, определение количества товаров, попавших в сегмент AX
- Определение числа клиентов, которые потратили в торговой сети 100 рублей и менее за все время, а также средней стоимости покупок у нововизитника
- Определение числа клиентов, совершивших более 50 визитов за все время
- Количество товаров в продажах со средней ценой 10 рублей и менее
- Определение списка товаров, у которых изменился ABC-XYZ сегмент сравнивая последние 6 месяцев и весь период

Тема 4. Теория вероятности и математическая статистика в аналитике данных

Предмет теории вероятностей. Пространство элементарных исходов. Операции над событиями и их свойства. Классическое определение вероятности. Правила и формулы комбинаторики. Подсчет классической вероятности с помощью правил комбинаторики. Статистическая и геометрическая вероятности.

Теоремы о произведении и сумме событий. Формула Бернулли. Полная вероятность. Формула Байеса. Понятие случайной величины. Определение и примеры дискретной случайной величины. Арифметические операции двух случайных величин. Числовые характеристики дискретной случайной величины. Числовые характеристики некоторых дискретных случайных величин. Непрерывные случайные величины. Числовые характеристики непрерывных случайных величин. Основные распределения непрерывных случайных величин. Функция распределения многомерной случайной величины. Двумерное дискретное распределение. Условное математическое ожидание в условных законах распределения. Двумерная непрерывная случайная величина. Многомерное нормальное распределение. Закон больших чисел. Центральная предельная теорема. Выборочный метод математической статистики. Применение математической статистики. Вариационные ряды и

их характеристики. Оценивание распределения случайных величин. Свойства статистических оценок. Общая схема проверки статистических гипотез. Проверка гипотезы о виде распределения случайной величины. Эвристические методы: метод коллективного блокнота, мозговой штурм, кейс-метод. Традиционные методы: метод сравнения, метод относительных и сравнительных величин, графический метод, метод группировки. Детерминированный факторный анализ: метод цепных подстановок, метод абсолютных разниц, метод относительных разниц. Матричные методы бизнес-анализа информации. Анализируются информационные потребности компаний и организаций, приводится классификация информации с точки зрения обработки с помощью статистических методов и моделей с использованием ИТ. Рассматриваются основные методы анализа количественной и качественной информации, методы группировки данных, проверки статистических гипотез, примеры методов и моделей анализа качественных переменных, условия применимости структурного анализа к экономической информации, приводятся примеры, рассматривается ряд алгоритмов структурного анализа экономической информации.

Рассматривается практический кейс «Визуализация OLAP-кубов»

Описание:

На основе дата-сетов Transactions_diy.lgd, БАНКИ рф.xlsx и Справочник городов.xlsx необходимо выполнить следующие практические задания:

Построить OLAP кубы активов банков по федеральным округам и регионам РФ и определить:

- ТОП-10 банков по кредитам физлицам в городах с населением выше 500 тыс. человек
- ТОП-10 городов по показателю “Число банков на 100 тысяч населения” в городах с населением выше 100 тыс. человек
- Создать отчет по распределению средств физлиц по городам-миллионникам и банкам

Тема 5. Обзор языковых, программных, инструментальных средств аналитики данных

Виды аналитики. Продуктовая аналитика. Маркетинговая аналитика. BI-аналитика (Business Intelligence-аналитика).

Инструменты аналитики: Для сбора и хранения данных. Для анализа данных. Jupyter Notebook. Для визуализации данных. Power Point, Miro, Tableau, Power BI. Для прогнозирования данных: языки программирования Python, R и другие. Инструменты и программы для аналитики данных бесплатные и коммерческие. Топ-10 основных инструментов в работе аналитика. SQL (Structured Query Language) как специальный язык, который формирует запросы к базе данных по определённой структуре. SQL и разновидности, система управления базами данных MySQL, PostgreSQL.

Python и библиотеки: Pandas как простая библиотека, которая помогает анализировать неполные и неупорядоченные данные, Matplotlib. Библиотека для визуализации, которая позволяет строить разные графики: от простых до гистограмм и круговых диаграмм. NumPy. Библиотека для обработки больших массивов данных и облегчения математических операций с массивами. R - язык программирования и конкурент Python. Jupyter Notebook- универсальный инструмент аналитика данных. Jupyter как бесплатная интерактивная среда для языка Python. DBeaver - универсальный инструмент для работы с различными базами данных (MySQL, PostgreSQL, Oracle и др). Talend как ETL-инструмент, который упрощает и оптимизирует процесс интеграции данных. ETL-технологии (Extract, Transform, Load) — «извлечение, преобразование и загрузка». Tableau, Power BI, инструмент Microsoft — Power BI. Visual Studio Code. Excel, аналитика с помощью таблиц Microsoft. SAS как комплексное решение, которое под одной оболочкой объединяет все технологии.

Рассматривается практический кейс «Скользящее окно. Квантование»

Описание:

Изучается применение скользящего окна при решении задачи построения прогнозного спроса на товарную группу по “наивной” модели скользящего среднего (величина спроса равна средним продажам в штуках за последние три месяца), практика расчет ошибки MAPE

На основе дата-сета Transactions_diy.lgd определить:

- значение MAPE группы “Освещение”
 - количество товаров группы, которое будет продано с самой минимальной ошибкой MAPE согласно модели прогноза по среднему
- Визуализировав данные методом квантования:
- определить количество клиентов, которые попали в последний интервал при разбиении
 - рассчитать для каждого клиента его 2 “любимые” с помощью функции CumulativeSum()
 - рассчитать число клиентов с Любимой группой 1 “Освещение”, в том числе производя расчёт не по сумме, а по количеству товара
 - рассчитать число клиентов, у которых нет Любимой группы 2

Тема 6. Обзор методик обнаружения новых знаний, работа с большими данными

Обсуждаются отличия Data Mining от классических статистических методов анализа и OLAP-систем, рассматриваются типы закономерностей, выявляемых Data Mining (ассоциация, классификация, последовательность, кластеризация, прогнозирование). Описывается сфера применения Data Mining. Вводится понятие Web Mining. Подробно рассматриваются методы Data Mining: нейронные сети, деревья решений, методы ограниченного

перебора, генетические алгоритмы, эволюционное программирование, кластерные модели, комбинированные методы. Знакомство с каждым методом проиллюстрировано решением практической задачи с помощью инструментального средства, использующего технологию Data Mining. Излагаются основные концепции хранилищ данных и места Data Mining в их архитектуре. Вводятся понятия OLTP, OLAP, ROLAP, MOLAP. Обсуждается процесс анализа данных с помощью технологии Data Mining. Подробно рассматриваются этапы этого процесса. Анализируется рынок аналитического программного обеспечения, описываются продукты от ведущих производителей Data Mining, обсуждаются их возможности.

Рассматривается практический кейс «Прогнозирование вероятности отказа клиента от заказа»

Описание: необходимо определить вероятность отказа клиента от заказа для интернет-магазина utkonos.ru

Актуальность задачи: отказ клиента от заказа во время доставки – актуальная проблема для компании в электронной торговле, ведущая за собой убытки в логистике, убытки из-за скоропортящихся товаров и прочие издержки

Дата-сет – 9 млн. транзакций, 354,8 тыс. заказов, 31,2 тыс. клиентов за период 2017-2018 год.

Этапы решения задачи: разведочный анализ, очистка данных – выдвижение гипотез и генерация признаков – проверка гипотез и отбор признаков, построение тривиальной модели, построение многофакторной модели – анализ результатов.

Проверка гипотез на основе WoE- и IV-анализа

Содержание семинаров, практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено занятие семинарского типа	Формы и методы проведения
1	Краткое описание: кто такой технический аналитик? Особенности профессии. Плюсы и минусы профессии технического аналитика. Важные личные качества	Решение практических заданий
2	Методы и средства моделирования бизнес-процессов Описание наиболее распространенных методов: - о методе функционального моделирования SADT (IDEF0); - о методе моделирования процессов IDEF3;	Решение практических заданий

	<ul style="list-style-type: none"> - моделирование потоков данных DFD; - о методе ARIS; - о методе Ericsson-Penker; - о методе моделирования, используемый в технологии Rational Unified Process. 	
3	<p>О предиктивной (прогнозной, предсказательной) аналитике</p> <p>О предписывающем анализе данных</p>	Решение практических заданий
4	О теории вероятности и математической статистике в аналитике данных (расчёт характеристики случайных величин, формулирование статистических гипотез и их проверка)	Решение практических заданий
5	О средствах аналитики данных	Решение практических заданий
6	О методиках обнаружения новых знаний	Решение практических заданий

Содержание самостоятельной работы слушателей по модулю

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
1.1.	Тема 1. Профессия «Технический аналитик (Technical Analyst Data Science)»: область деятельности, необходимые хард и софт скилы	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.2.	Тема 2. Основные понятия и методы моделирования бизнес процессов	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

1.3.	Тема 3. Бизнес-анализ: предиктивная аналитика, предписывающий анализ	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.4.	Тема 4. Теория вероятности и математическая статистика в аналитике данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.5.	Тема 5. Обзор языковых, программных, инструментальных средств аналитики данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.6.	Тема 6. Обзор методик обнаружения новых знаний, работа с большими данными	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации

1. Роль информации в экономике
2. Классификация документированных источников знания
3. Законодательное регулирование информационной деятельности
4. Информационные потребности компаний и организаций
5. Структура информационного обеспечения компании
6. Процесс управления информационными ресурсами
7. Информационные потребности при принятии управленческих решений
8. Мировой рынок информации
9. Информационная среда интернет
10. Статистические информационные ресурсы
11. Особенности работы с интернет-ресурсами
12. Технологии управления информацией и знаниями в компании
13. Общая технологическая схема управления информацией
14. Технологии отбора источников информации
15. Подходы к оценке информации
16. Методы и модели экономического анализа информации
17. Методы и модели анализа финансовой информации
18. Применение финансовой математики для бизнес-анализа информации
19. Графовые модели анализа финансовой информации
20. Бенчмарк-анализ экономической информации

Перечень основной и дополнительной учебной литературы

Нормативно-правовые акты:

1. Национальная программа "Цифровая экономика Российской Федерации", утв. распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р

2. Федеральная проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 № 17).

Основная литература:

3. Сурова Н.Ю. Искусственный интеллект. Монография. Гриф НИИ образования и науки. Гриф МУМЦ "Профессиональный учебник" Гриф МНИЦ Судебной экспертизы и исследований – 2 – е издание М.: Изд-во ЮНИТИ, 2021, 360 с.

4. Бенгфорт, Б. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка / Б. Бенгфорт. — СПб.: Питер, 2021. — 368 с.

5. Бендерская, О. Б. Бизнес-аналитика: учебное пособие / О. Б. Бендерская Бизнес-аналитика, Весь срок охраны авторского права Электрон. дан. (1 файл) Белгород: Белгородский государственный технологический университет им. В.Г. Шухова, ЭБС АСВ, 2020. - 162 с.

6. Основы бизнес-анализа. Учебное пособие / Под ред. Бариленко В.И. - М.: КноРус, 2020. - 56 с.

7. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2021-336 с.: ил. (Серия «Библиотека программиста»).

8. Форман Дж. Много цифр: Анализ больших данных при помощи Excel / Дж. Форман. — М.: Альпина Паблишер, 2020. — 461 с.

9. Data Science. Наука о данных с нуля. / Билл Фрэнкс.; пер. с англ. Евстигнеева И.В. – М.: Издательство «Альпина Паблишер». – 2020. – 320 с.

10. Набатова Д. С. Математические и инструментальные методы поддержки принятия решений: учебник и практикум для вузов / Д. С. Набатова. — Москва: Издательство Юрайт, 2022. — 292 с. — (Высшее образование). — ISBN 978-5-534-02699-3.

11. Курносков Ю.В. «Азбука аналитики», Издательство «Концептуал», 2022 -240 с.

12. Б. Марр «Ключевые инструменты бизнес-аналитики» / пер с англ. Егоров В. Н., Издательство «Лаборатория знаний», 2022 – 339 с.

Дополнительная литература:

13. Вигерс Карл, Битти Джой. Разработка требований к программному обеспечению. 4-е изд., дополненное / Пер. с англ. — М.: Издательство «Русская редакция»; СПб. : БХВ-Петербург, 2020. — 736 стр.: ил.

14. Винстон, У. Бизнес-моделирование и анализ данных. Решение актуальных задач с помощью Microsoft Excel / У. Винстон. - СПб.: Питер, 2021. - 320 с.

15. Воловиков, Б.П. Стратегическое бизнес-планирование на промышленном предприятии с применением динамических моделей и сценарного анализа: Монография / Б.П. Воловиков. - М.: Инфра-М, 2021. - 320 с.

16. Еремеева, Н.В. Планирование и анализ бизнес-процессов на основе построения моделей управления конкурентоспособности продукции / Н.В. Еремеева. - М.: Русайнс, 2022. - 16 с.

17. Казакова, Н.А. Экономический анализ в оценке бизнеса и управлении инвестиционной привлекательностью компании / Н.А. Казакова. - М.: Финансы и статистика, 2022. - 240 с.

18. Казакова, Н.А. Экономический анализ в оценке бизнеса и управлении инвестиционной привлекательностью компании: Учебное пособие / Н.А. Казакова. - М.: Финансы и статистика, 1. - 240 с.

19. Карлберг, К. Бизнес-анализ с использованием Excel / К. Карлберг. - М.: Диалектика, 2020. - 576 с.

20. Кулаичев А.П. Методы и средства комплексного анализа данных: Учебное пособие / А.П. Кулаичев. — М.: Форум, 2022. — 160 с.
Кэхилл, М. Инвестиционный анализ и оценка бизнеса: Учебное пособие / М. Кэхилл. - М.: ДиС, 2022. - 432 с.

21. Лацис, А.О. Параллельная обработка данных / А.О. Лацис. - М.: Academia, 2021. - 456 с.

22. Нархид Н. Apache Kafka. Поточковая обработка и анализ данных / Н. Нархид. — СПб.: Питер, 2020. — 320 с.

23. Ниворожкина Л.И. Статистические методы анализа данных: Учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга. — М.: Риор, 2022. — 320 с.

24. Орлова, Е.Р. Бизнес-план: Методика составления и анализ типовых ошибок / Е.Р. Орлова. - М.: Омега-Л, 2022. - 168 с.

25. Фляйшер, К. Стратегический и конкурентный анализ: Методы и средства конкурентного анализа в бизнесе / К. Фляйшер, Б. Бенсуссан; Пер. с англ. Д.П. Конькова - М.: БИНОМ. ЛЗ, 2022. - 541 с.

Интернет-ресурсы

1. <https://biconsult.ru/products/uchebnoe-posobie-po-tableau> (дата обращения: 22.03.2021). Учебник по Tableau - система бизнес-анализа (BI).

2. <https://biconsult.ru/products/uchebnoe-posobie-po-tableau> (дата обращения: 22.03.2021). Учебник по Tableau - система бизнес-анализа (BI).

3. [http://www.tadviser.ru/index.php/Статья:Большие_данные_\(Big_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data))
Статья «Большие данные (Big Data)» [Электронный ресурс]

4. <https://powerbi.microsoft.com/ru-ru/desktop/> - страница загрузки платформы Power BI Desktop компании Microsoft

5. <https://public.tableau.com/en-us/s/download> - страница загрузки платформы Tableau Public

6. <https://loginom.ru/download> - страница загрузки платформы Loginom компании BaseGroup Labs

7. <https://www.knime.com/knime-analytics-platform> - страница загрузки платформы Knime Analytics Platform

8. <https://cloud.yandex.ru/services/datalens> – сервис визуализации и анализа данных Яндекс

Описание системы оценки качества освоения дисциплины

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в ходе проведения практических занятий.

Пример практических заданий:

Задание 1. Данные по компании:

Денежные средства и краткосрочные финансовые вложения	100 ден. ед.
Основные средства и другие внеоборотные активы	283,5 ден. ед.
Выручка от продаж за год	1 000 ден. ед.
Прибыль за год	50 ден. ед.
Коэффициент промежуточного покрытия	2
Коэффициент текущей ликвидности	3
Оборачиваемость дебиторской задолженности	40 дней
Рентабельность собственных источников	12%

Определить:

1. Величину средней дебиторской задолженности
2. Краткосрочные обязательства
3. Оборотные средства
4. Общую сумму активов
5. Рентабельность активов
6. Собственный капитал

Задание 2. Сдельная зарплата каждого из 5 рабочих одной квалификации составила соответственно 2, 1, 3, 8, 1 д.е. Эти данные использованы для изучения показателя - сдельной зарплате одного рабочего. Если к среднему значению по выборке прибавить моду, а затем вычесть медиану выборки, то получится?

Задание 3. Компанию по прокату автомобилей интересует зависимость между пробегом автомобилей X (тыс. км) и стоимостью технического обслуживания Y тыс.руб. Для выяснения характера связи было отобрано 30 автомобилей и получены оценки:

$\sum Y_i = 30$; $\sum X_i = 90$; $\sum X_i Y_i = 135$; $\sum Y_i^2 = 330$; $\sum X_i^2 = 360$. Оценить тесноту линейной связи между переменными с помощью выборочного коэффициента корреляции.

Задание 3. Каждый квартал на банковский счет преподавателя организация перечисляет 1000 руб., на которые банк начисляет каждые полгода 10% по схеме сложных процентов. Сколько будет на счете через 3 года и 3 недели?

Промежуточный контроль: тест. Слушателю предлагается решить тестовые задания.

Пример Тестовых заданий

Аналитик это ...

- а) специалист в области анализа и моделирования
- б) специалист в предметной области;
- в) человек, решающий определенные задачи;
- г) человек, который имеет опыт в программировании.

Эксперт это ...

- а) специалист в области анализа и моделирование;
- б) специалист в предметной области;
- в) человек, решать определенные задачи;
- г) человек, который имеет опыт в программировании.

Задача классификации сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Задача регрессии сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристикам объекта значения некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Задача кластеризации заключается в ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;

в) определение по известным характеристикам объекта значения некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Целью поиска ассоциативных правил является ...

а) нахождения частых зависимостей между объектами или событиями;

б) определения класса объекта по его характеристиками;

в) определение по известным характеристикам объекта значение некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

До предполагаемых моделей относятся такие модели данных:

а) модели классификации и последовательностей;

б) регрессивные, кластеризации, исключений, итоговые и ассоциации;

в) классификации, кластеризации, исключений, итоговые и ассоциации;

г) модели классификации, последовательностей и исключений.

К описательным моделям относятся следующие модели данных:

а) модели классификации и последовательностей;

б) регрессивные, кластеризации, исключений, итоговые и ассоциации;

в) классификации, кластеризации, исключений, итоговые и ассоциации;

г) модели классификации, последовательностей и исключений.

Модели классификации описывают ...

а) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;

б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;

в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

Модели последовательностей описывают ...

а) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;

б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;

в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

Экономический анализ изучает:

- а) все хозяйственные процессы в организации, рассматриваемые с точки зрения целесообразности
- б) финансовые и трудовые ресурсы с точки зрения их оптимизации
- в) материальные ресурсы, деятельность хозяйствующих субъектов, результат деятельности

Экономическая информация – это:

- а) система сведений о всех видах экономической деятельности организаций, сопровождает процессы производства, распределения, обмена и потребления материальных благ и услуг
- б) система сведений о видах экономической деятельности организаций, таких как производство и потребление материальных благ и услуг
- в) система сведений о производственно-сбытовой, снабженческой и финансовой деятельности, рассматриваемой с точки зрения эффективности.

По степени воздействия на изучаемый объект показатели разделяются на ...

- а) обобщающие и частные
- б) результативные и факторные
- в) учетные и отчетные

В следующем списке показателей укажите количественный показатель:

- а) себестоимость продукции,
- б) производительность труда
- в) объем выпуска продукции

Как связаны между собой методология и методика?

- а) методология часть методик
- б) методика часть методологии
- в) это одно и то же явление

Какой из способов, измеряющих влияние факторов, более точно отражает исследуемое

- а) метод цепных подстановок
- б) способ абсолютных разниц
- в) интегральный метод

Отношение чистой прибыли к собственному капиталу показывает...

- 1) ROA
- 2) ROE
- 3) ROI
- 4) RonA

По алгоритму Форда-Фалкерсона получен граф, в котором вершину t нельзя пометить. Что это означает?

- А) получен допустимый поток
- Б) получен минимальный разрез
- В) нет решения

Порядок проведения: тестирование проводится с личного компьютера, 10 тестовых вопросов по всем темам дисциплины, 60 мин, количество попыток – 3.

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Промежуточная аттестация проводится на основании балльно-рейтинговой системы. В процессе обучения учебные действия слушателя оцениваются по шкале:

Вид учебного действия	Максимальное количество баллов
Академическая активность (посещаемость, активность на семинарах и т.д.)	15
выполнение всех практических заданий	20
текущая аттестация (по всем темам)	25
промежуточный контроль по дисциплине	30
ВСЕГО	100

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Рабочая программа модуля 2

«Современные технологии хранения и обработки данных»

Цель освоения модуля 2 – приобретение слушателями знаний по современным технологиям хранения и обработки данных, приобрести умения работы со статистическими методами анализа, обработки данных и нейронными сетями, а также овладеть навыками обработки данных с использованием табличных процессоров, а также создания нейронных сетей.

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 2:

- ПК-1. Применяет стандарты и методики бизнес-моделирования.
- ПК-2. Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.

Планируемые результаты обучения по модулю 2

По итогам освоения модуля слушатели должны:

Знать:

- Технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памяти;
- Статистический анализ: метод многовариантного тестирования, корреляционный анализ, регрессионный анализ;
- Статистические методы: параметрические, непараметрические, управляемые, неуправляемые, полууправляемые, кластеризация. Алгоритмы машинного обучения: обучение с учителем, обучение без учителя, полууправляемое обучение, обучение с подкреплением. Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация;
- Принципы работы нейронных сетей.

Уметь:

- Использовать технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памяти;
- Проводить сравнительный анализ методов и инструментальных средств анализа больших данных;
- Планировать аналитические работы с использованием технологий больших данных
- Проводить анализ больших данных
- Осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных;
- Создавать нейронные сети для решения задач.

Владеть:

– Навыками использования технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памяти;

– Навыками использования технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, планирование экспериментов, карты контроля качества;

– Навыками работы с библиотекой TensorFlow.

Учебно-тематический план модуля 2

«Технологии хранения и обработки данных»

№ п/п	Наименование дисциплины, модуля, темы	Трудоемкость		В том числе			Самостоятельная	Форма контроля
		В зачетных единицах	В часах	Всего	Контактная работа ⁴			
					из них	Лекции		
1	2	3	4	5	6	7	8	9
1	МОДУЛЬ 2							
2	Тема 1. Современные технологии хранения и обработки данных		18	12	6	6	6	
3	Тема 2. Статистические методы анализа и обработки данных		18	12	6	6	6	
4	Тема 3. Обработка данных с использованием табличных процессоров. Нейронные сети		22	15	8	7	7	
5.	Промежуточная аттестация		2	2	-	2		-
6.	Итого по модулю 1		60	39	20	19	19	2, тест

⁴ С применением дистанционных образовательных технологий и электронного обучения

Содержание модуля 2

«Технологии хранения и обработки данных»

Тема 1. Современные технологии хранения и обработки данных

Данные и информация. Основные понятия – данные, информация, метаинформация, знания. Виды информации. Обработка данных и ее виды. Основные процедуры обработки данных. Создание, модификация данных, контроль, безопасность и целостность данных, поиск информации, поддержка принятия решения, создание документов, сводок, отчетов, преобразование информации. Модели процессов обработки данных. Задачи обработки данных. Инструменты обработки данных и визуализация. Первичная обработка и анализ данных. Основы баз данных. Проектирование данных. Запросы на языке SQL. Объекты базы данных. Задачи обработки данных различных типов. Прикладные области обработки данных. Оцифровка сигналов. Теорема Котельникова. Сжатие информации, алгоритмы сжатия. Параллельная и распределенная обработка информации Многопоточные приложения. Распределенная и параллельная обработка запросов, распределенная обработка транзакций. Хранилища данных. Технология OLAP. Понятие гиперкуба. Тест FASMI. Структура OLAP куба. Компоненты OLAP. Архитектура OLAP. Данные в ХД: детализированные и агрегированные данные, метаданные. Технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенная и параллельная обработка данных, вычисления в оперативной памяти.

Рассматривается практический кейс 1 «Сегментация тканей головного мозга с использованием алгоритмов кластеризации в Matlab»

Описание:

На практике изучаются методы работы с данными в Excel, построение гистограмм распределения, основные понятия и характеристики математической статистики, корреляционный анализ, регрессионный анализ.

Используемые дата-сеты: Salaries.xlsx, natural_gas_price_daily.xlsx, CompanyABCprofit.xlsx, brent_oil-daily.xlsx

Рассматривается практический кейс 2 «Сегментация тканей головного мозга с использованием алгоритмов кластеризации в Matlab»

Описание:

Основная задача данного задания состоит в том, чтобы с помощью алгоритма кластеризации (без учителя) сегментировать мультимодальные изображения МРТ головного мозга. Для каждой из следующих видов тканей должны быть назначены разные кластеры /метки: серое вещество (GM), белое

вещество (WM) и спинномозговая жидкость (CSF). Подобный вид визуализации тканей мозга полезен в исследованиях, связанных с болезнью Альцгеймера или измерения потери серого вещества. Исходные данные получены из репозитория Brainweb (<http://brainweb.bic.mni.mcgill.ca/brainweb/>). Они представляют из себя трёхмерные мультимодальные МРТ головного мозга с размер вокселя 1x1x1 мм. и 5% шумом (без моделирования неоднородности интенсивности). Набор данных имеет размер 181x217x181.

Для чтения МРТ данных используется функция `Read_BrainWebData()`. Для решения задачи кластеризации использовать алгоритм k-means.

Данная работа выполняется в среде MATLAB с помощью встроенного пакета «Image Processing Toolbox».

Выходными аргументами являются – отсегментированное изображение, финальная точность сегментации, а также нормализованная матрица ошибок (confusion matrix).

Тема 2. Статистические методы анализа и обработки данных

Основы науки о данных (Data Science), роль математических методов и информационных технологий в проведении статистического анализа, представление о статистике, больших данных и основных методах анализа данных. Статистика. Выборочный метод. Вариационные ряды, динамические ряды и их обработка. Анализ взаимосвязей. Многомерный статистический анализ. Математическое моделирование. Классификация и регрессия. Постановка задачи. Классификационные правила. Методы построения правил классификации. Алгоритм построения правил. Метод Naïve Bayes. Деревья решений. Методы построения деревьев решений. Математические функции. Методы построения математических функций. Прогнозирование временных рядов. Методы прогнозирования временных рядов. Алгоритм Apriori. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации. Представление результатов. Алгоритмы кластеризации. Иерархические алгоритмы. Неиерархические алгоритмы. Адаптивные методы кластеризации.

Рассматривается практический кейс 2 «Сегментация тканей головного мозга с использованием алгоритмов кластеризации в Matlab»

Описание: цель кейса – ознакомиться с алгоритмами машинного и глубокого обучения и их применением для решения задачи классификации медицинских изображений.

Основная задача состоит в том, чтобы с помощью алгоритмов машинного и глубокого обучения (обучение с учителем) классифицировать различные ткани лёгких. Данная задача относится к решению проблемы мульти-классовой классификации с четырьмя разными типами ткани: здоровая ткань (HT-0), парасептальная эмфизема (PS-1), панлобулярная эмфизема (PS-1) и центрилобулярная эмфизема (CL-3).

Набор данных для решения поставленной задачи состоит из 267 изображений компьютерной томографии лёгких. Для разметки данных профессиональный пульмонолог вручную выбрал регионы, соответствующие четырём классам, а именно 370 регионов для здоровой ткани, 184 региона для парасептальной эмфиземы, 148 регионов для панлобулярной эмфиземы и 287 регионов для центрилобулярной эмфиземы. Общее количество сэмплов в данном наборе данных – 989. Каждый сэмпл представляет из себя изображение 31*31 пиксель. Примеры сэмплов отражены ниже на рисунке 1.

Тема 3. Обработка данных с использованием табличных процессоров. Нейронные сети

Табличный процессор как средство обработки данных, структура электронной таблицы и типы данных; встроенные функции; деловая графика; построение графиков функций; применение сортировки и фильтров; поиск решения и подбор параметр, универсальные табличные процессоры (MEExcel, Lotus, QuatroPro); специализированные бухгалтерские программы (1-С, Инфо-Бухгалтер); специализированные банковские программы (внутрибанковских и межбанковских расчетов); специализированные программы финансово-экономического анализа и планирования (ПАРУС, ГАЛАКТИКА). Функции табличных процессоров. Нейронные сети: общая архитектура. Многослойные сети. Обратное распространение ошибки. Библиотека TensorFlow.

Рассматривается практический кейс 1 «Многослойный перцептрон»
Применение нейронной сети для классификации изображений

Содержание семинаров, практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено занятие семинарского типа	Формы и методы проведения
1	О современных технологиях хранения и обработки данных (применение, организация системы хранения данных, варианты подключений, основные системы)	Решение практических заданий
2	О создании системы хранения данных	Решение практических заданий
3	О методах статистического анализа данных	Решение практических заданий
4	Об обработке табличных данных	Решение практических заданий

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
1.1.	Тема 1. Современные технологии хранения и обработки данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.2.	Тема 2. Статистические методы анализа и обработки данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.3.	Тема 3. Обработка данных с использованием табличных процессоров. Нейронные сети	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации

1. Понятие информационные технологии.
2. Понятие количества информации. Методы оценки
3. Понятие информационной технологии. Этапы развития
4. Основные функции базовой информационной технологии
5. Этапы развития (эволюция) информационных технологий
6. Нарисуйте схему преобразования «информация - данные»
7. Уровни базовой информационной технологии
8. Нарисуйте схему концептуальной модели базовой информационной технологии
9. Определите термины информационный процесс, информационная процедура, информационная операция.
10. Чем отличаются процессы преобразования информации и процессы преобразования данных?
11. В чем состоят процессы получения, подготовки и ввода информации?
12. В чем смысл процесса обработки данных и его процедур?
13. Каковы функции процесса и процедур обмена данными?
14. Для чего используются процесс и процедуры накопления данных?
15. Опишите назначение и суть процесса и процедур представления знаний.

16. Что такое логический уровень информационной технологии, для чего необходимо его рассмотрение?
17. Нарисуйте схему состава моделей базовой информационной технологии и объясните назначение и связи каждой модели.
18. Каким образом информационная технология отображается на физическом уровне?
19. Нарисуйте схему состава и взаимосвязей подсистем базовой информационной технологии и поясните, на каких аппаратно-программных средствах они реализуются.
20. Какова последовательность преобразования информации в данные?
21. Какие методы контроля применяются в процессе преобразования информации в данные?
22. Основные этапы обработки экономической информации
23. Способы обработки экономической информации. Достоинства и недостатки
24. Организации пакетного режима обработки данных
25. Характеристика диалогового режима обработки данных.
26. Характеристика технологий обработки данных в режиме реального времени
27. Характеристика пакетного режима обработки данных.
28. В чем состоит суть процедуры преобразования данных и как она реализуется в ЭВМ?
29. Для чего служит процедура отображения данных и какие операции ее реализуют?
30. Что служит теоретической базой для создания моделей компьютерной графики?
31. Опишите два основных метода получения графического изображения на экране монитора.
32. На каких аппаратно-программных средствах реализуется информационный процесс обработки данных.
33. Процедуры хранения данных. Базы данных.
34. Средства создания электронного документа. Текстовые редакторы.
35. Обработка и анализ экономической информации с помощью электронных таблиц EXCEL.
36. Средства разработки мультимедийных презентаций. Работа с Power Point
37. Организация и ведение баз данных средствами СУБД MS ACCESS
38. Основные виды и процедуры обработки информации. Централизованная, децентрализованная, распределенная и интегрированная обработка.
39. Виды операций обработки данных - сбор, анализ, поиск, представление, хранение.
40. Структуры данных. Линейные структуры данных. Нелинейные структуры данных.
41. Модели и методы решения задач обработки информации.

42. Современные подходы к анализу данных.
43. Базовая терминология анализа данных, понятие модели и моделирования.
44. Классификация программных продуктов для создания аналитических решений. Характеристики аналитических платформ.
45. Основные задачи консолидации данных, Обобщенная схема процесса консолидации.
46. Цели трансформации и ее роль в процессе обработки данных. Основные методы трансформации.
47. Основные методы нормализации данных. Нормализация с помощью поэлементных преобразований.
48. Цели и задачи визуализации данных. Группы методов визуализации.
49. Концепция управления качеством данных. Уровни качества данных. Оценка качества данных.
50. Алгоритмы построения и анализа данных. Алгоритмы поиска.
51. Алгоритмы сортировки.
52. Современные технологии хранения данных. Классификация запоминающих устройств.
53. Сетевые технологии обработки информации.
54. Классификация сетевых технологий.
55. Характеристика беспроводных и кабельных сетевых технологий.
56. Принципы формирования сетевых информационных хранилищ.
57. Распределение прав доступа. Исключение несанкционированного изменения информации.
58. Что такое нейронная сеть. Однослойные и многослойные нейронные сети.
59. Обратное распространение ошибки. Особенности и возможности библиотеки TensorFlow/

Перечень основной и дополнительной учебной литературы

Нормативно-правовые акты:

1. Национальная программа "Цифровая экономика Российской Федерации", утв. распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р
2. Федеральной проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 N 17).

Основная литература:

3. Сурова Н.Ю. Искусственный интеллект. Монография. Гриф НИИ образования и науки. Гриф МУМЦ "Профессиональный учебник" Гриф

МНИЦ Судебной экспертизы и исследований – 2 – е издание М.: Изд-во ЮНИТИ, 2021, 360 с.

4. Бенгфорт, Б. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка / Б. Бенгфорт. — СПб.: Питер, 2020. — 368 с.

5. Бендерская, О. Б. Бизнес-аналитика: учебное пособие / О. Б. Бендерская Бизнес-аналитика, Весь срок охраны авторского права Электрон. дан. (1 файл) Белгород: Белгородский государственный технологический университет им. В.Г. Шухова, ЭБС АСВ, 2017. - 162 с.

6. Бурнаева, Э., Г. Обработка и представление данных в MS Excel: Учебное пособие / Э.Г. Бурнаева, С.Н. Леора. - СПб.: Лань, 2021. - 160 с.

7. Волкова, П.А. Статистическая обработка данных в учебно-исследовательских работах: Учебное пособие / П.А. Волкова, А.Б. Шипунов. - М.: Форум, 2017. - 832 с.

8. Основы бизнес-анализа. Учебное пособие / Под ред. Бариленко В.И. - М.: КноРус, 2020. - 56 с.

9. Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2017-336 с.: ил. (Серия «Библиотека программиста»).

10. Форман Дж. Много цифр: Анализ больших данных при помощи Excel / Дж. Форман. — М.: Альпина Паблишер, 2020. — 461 с.

11. Data Science. Наука о данных с нуля. / Билл Фрэнкс.; пер. с англ. Евстигнеева И.В. – М.: Издательство «Альпина Паблишер». – 2022. – 320 с.

12. Набатова Д. С. Математические и инструментальные методы поддержки принятия решений: учебник и практикум для вузов / Д. С. Набатова. — Москва: Издательство Юрайт, 2022. — 292 с. — (Высшее образование). — ISBN 978-5-534-02699-3.

13. Курносов Ю.В. «Азбука аналитики», Издательство «Концептуал», 2022 -240 с.

14. Б. Марр «Ключевые инструменты бизнес-аналитики» / пер с англ. Егоров В. Н., Издательство «Лаборатория знаний», 2022 – 339 с.

Дополнительная литература:

15. Аббакумов, В. Бизнес-анализ информации. Статистические методы: Учебник / В. Аббакумов, Т. Лезина. - М.: Экономика, 2022. - 374 с.

16. Аббакумов, В., Л. Бизнес- анализ информации. Статистические методы / В.Л. Аббакумов. - М.: Экономика, 2022. - 374 с.

17. Барков, С.А. Бизнес в литературе: социологический анализ / С.А. Барков, В.И. Зубков. - М.: Аккадемический проект, 1. - 253 с.

18. Брускин, С.Н. Интеллектуальный анализ динамики бизнес-систем / С.Н. Брускин. - М.: Инфра-М, 2022. - 320 с.

19. Брускин, С.Н. Интеллектуальный анализ динамики бизнес-систем: Учебник / С.Н. Брускин. - М.: Инфра-М, 2022. - 320 с.

20. Винстон, У. Бизнес-моделирование и анализ данных. Решение актуальных задач с помощью Microsoft Excel / У. Винстон. - СПб.: Питер, 2022. - 320 с.
21. Воловиков, Б.П. Стратегическое бизнес-планирование на промышленном предприятии с применением динамических моделей и сценарного анализа: Монография / Б.П. Воловиков. - М.: Инфра-М, 2021. - 320 с.
22. Еремеева, Н.В. Планирование и анализ бизнес-процессов на основе построения моделей управления конкурентоспособности продукции / Н.В. Еремеева. - М.: Русайнс, 2022. - 16 с.
23. Казакова, Н.А. Экономический анализ в оценке бизнеса и управлении инвестиционной привлекательностью компании / Н.А. Казакова. - М.: Финансы и статистика, 2022. - 240 с.
24. Казакова, Н.А. Экономический анализ в оценке бизнеса: учебно-практическое пособие / Н.А. Казакова. - М.: ДиС, 2022. - 288 с.
25. Казакова, Н.А. Экономический анализ в оценке бизнеса и управлении инвестиционной привлекательностью компании: Учебное пособие / Н.А. Казакова. - М.: Финансы и статистика, Инфра-М, 2022. - 240 с.
26. Казакова, Н.А. Экономический анализ в оценке бизнеса и управлении инвестиционной привлекательностью компании: Учебное пособие / Н.А. Казакова. - М.: Финансы и статистика, 1. - 240 с.
27. Карзаева, Н.Н. Экономический анализ в оценке бизнеса и управлении инвестиционной привлекательностью компании: Учебное пособие / Кара-Ушанов В.Ю., - 2-е изд., стер. - Москва: Флинта, Изд-во Урал. ун-та, 2021. - 156 с.
28. Карлберг, К. Бизнес-анализ с использованием Excel / К. Карлберг. - М.: Диалектика, 2020. - 576 с.
29. Кулаичев А.П. Методы и средства комплексного анализа данных: Учебное пособие / А.П. Кулаичев. — М.: Форум, 2022. — 160 с.
30. Кэхилл, М. Инвестиционный анализ и оценка бизнеса: Учебное пособие / М. Кэхилл.. - М.: ДиС, 2022. - 432 с.
31. Лацис, А.О. Параллельная обработка данных / А.О. Лацис. - М.: Academia, 2021. - 456 с. SQL - язык реляционных баз данных: Учебное пособие / Н.Н. Карзаева. - М.: Финансы и статистика, 2022. - 240 с.
32. Нархид Н. Apache Kafka. Поточковая обработка и анализ данных / Н. Нархид. — СПб.: Питер, 2020. — 320 с.
33. Усенко, Л.Н. Бизнес-анализ деятельности организации: Учебник / Л.Н. Усенко, Ю.Г. Чернышева, Л.В. Гончарова и др. - М.: Альфа-М, 2021. - 512 с.
34. Фляйшер, К. Стратегический и конкурентный анализ: Методы и средства конкурентного анализа в бизнесе / К. Фляйшер, Б. Бенсуссан; Пер. с англ. Д.П. Конькова. - М.: БИНОМ. ЛЗ, 2022. - 541 с.
35. Вигерс Карл, Битти Джой. Разработка требований к программному обеспечению. 4-е изд., дополненное / Пер. с англ. — М.: Издательство «Русская редакция»; СПб. : БХВ-Петербург, 2020. — 736 стр.: ил.

Интернет-ресурсы

36. <http://h2o-release.s3.amazonaws.com/h2o/rel-zermelo/4/index.html> - страница загрузки платформы H2O
37. <https://cloud.google.com/bigquery/docs/sandbox> - страница BigQuery sandbox
38. [http://www.tadviser.ru/index.php/Статья:Большие_данные_\(Big_Dat\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Dat)) Статья «Большие данные (Big Data)» [Электронный ресурс]
39. <https://biconsult.ru/products/uchebное-posobie-po-tableau> (дата обращения: 22.03.2021). Учебник по Tableau - система бизнес-анализа (BI).
40. <https://biconsult.ru/products/uchebное-posobie-po-tableau> (дата обращения: 22.03.2021). Учебник по Tableau - система бизнес-анализа (BI).
41. <https://cloud.yandex.ru/services/datalens> – сервис визуализации и анализа данных Яндекс
42. <https://community.cloud.databricks.com/login.html> - страница регистрации Databricks Community Edition
43. <https://rapidminer.com/get-started/> - страница загрузки платформы RapidMiner
44. <https://loginom.ru/download> - страница загрузки платформы Loginom компании BaseGroup Labs
45. <https://powerbi.microsoft.com/ru-ru/desktop/> - страница загрузки платформы Power BI Desktop компании Microsoft
46. <https://public.tableau.com/en-us/s/download> - страница загрузки платформы Tableau Public
47. <https://urait.ru/bcode/413823> (дата обращения: 10.04.2022).. Зараменских Е. П. Основы бизнес-информатики: учебник и практикум для бакалавриата и магистратуры / Е. П. Зараменских. – Москва: Издательство Юрайт, 2022. — 407 с. — (Высшее образование). — ISBN 978-5-9916-8210-7. — Текст: электронный // Образовательная платформа Юрайт [сайт]
48. <https://www.knime.com/knime-analytics-platform> - страница загрузки платформы Knime Analytics Platform

Описание системы оценки качества освоения дисциплины

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в ходе проведения практических занятий.

Пример практических заданий:

Задание 1.

Напишите команду SELECT, которая бы вывела номер заказа, сумму и дату для всех строк из таблицы Orders.

Напишите запрос, который вывел бы все строки из таблицы Customers, с номером продавца = 1001.

Напишите запрос, который вывел бы таблицу со столбцами в следующем порядке: city, sname, snum, comm.

Напишите команду SELECT, которая вывела бы рейтинг (rating), сопровождаемый именем каждого покупателя в San Jose.

Напишите запрос, который вывел бы значения snum всех продавцов из таблицы Orders, без каких бы то ни было повторений.

Задание 2. Используя фрейм present, создайте график доли мальчиков с течением времени, и на основе графика определите, является ли следующее утверждение истинным или ложным: доля мальчиков, уменьшилась с течением времени. Совет: с помощью клавиш со стрелками вверх и вниз можно просматривать все предыдущие команды. Также можно получить доступ к истории команд, нажав на вкладку History в верхней правой панели. Это позволит сэкономить много времени в будущем.

Задание 3. Используя фрейм present, создайте график, который отображает соотношение мальчик-к-девочке за каждый год. Что вы видите? Опишите тенденцию. *Задание 3* Используя фрейм present, вычислите абсолютные различия между количеством мальчиков и девочек, родившихся в каждом году, и определите, в каком году была самая большая абсолютная разница в количествах новорожденных девочек и мальчиков?

Задание 4. Используя фрейм diamonds, определите все возможные сочетания качества огранки (cut), цвета (color) и чистоты (clarity) алмазов. Результат представьте в таблице со столбцами: No п/п, cut, color, clarity.

Задание 5. Используя фрейм diamonds, определите 100 самых больших алмазов, с указанием их цены (price), величины (carat) и размеров по трем измерениям (x, y, z).

Промежуточный контроль: тест. Слушателю предлагается решить тестовые задания.

Пример Тестовых заданий

Из представленных адресов широковещательным адресом класса C будет: а. 100.168.255.255. б. 129.168.253.255 с. 190.168.255.255. d. 224.168.253.255.

Сколько времени будет проходить передача файла размером 128 кбайт по сети, скорость которой составляет 128Кбит/с? а. 1 с. б. 10 с. с. 8 с. d. 16 с.

Какие устройства функционируют на канальном уровне модели OSI? а. коммутаторы. б. маршрутизаторы. с. повторители. d. мосты.

Для передачи файлов по сети используется протокол: а. SMTP. б. FTP. с. SMTP. d. HTTP.

Какой домен верхнего уровня означает "образовательный сайт"? а. com б. spb. с. edu d. gov

Наименьший номер порта, который может быть динамически назначен на конечный узел при его обращении к серверу, будет: а. 1. б. 255. с. 1023. d. 64.

Какая из приведенных схем соединения компьютеров представляет собой замкнутую цепочку? а. звезда. б. кольцо. с. логическая топология. d. шина.

Топология компьютерной сети, в которой все компьютеры сети присоединены к центральному узлу называется: а. логическая топология. б. звезда. с. кольцо. d. шина

Адрес 0005.A869.CD-F1 является: а. логическим б. почтовым адресом с. MAC-адресом d. номером порта

E-mail серверы для связи между собой используют следующий протокол: а. TFTP. б. FTP. с. HTTP. d. SMTP.

К технологиям локальных сетей относятся: а. PDH. б. SDH. с. Token Ring. d. Ethernet.

Для передачи файлов используется следующий протокол типа connectionless: а. IP. б. TCP. с. SMTP. d. TFTP.

Если приложение использует протокол, который при обмене данными не использует технологию скользящего окна или контроль потока, а для надежности должен полагаться на протоколы более высокого уровня, то для передачи используются следующие протокол и метод: а. IP. б. UDP, connection-oriented. с. TCP, connection-oriented. d. UDP, connectionless.

Порядок проведения: тестирование проводится с личного компьютера, 10 тестовых вопросов по всем темам дисциплины, 60 мин, количество попыток – 3.

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Промежуточная аттестация проводится на основании балльно-рейтинговой системы. В процессе обучения учебные действия слушателя оцениваются по шкале:

Вид учебного действия	Максимальное количество баллов
Академическая активность (посещаемость, активность на семинарах и т.д.)	15
выполнение всех практических заданий	20
текущая аттестация (по всем темам)	25
промежуточный контроль по дисциплине	30
ВСЕГО	100

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Рабочая программа модуля 3

«Языковые средства для аналитики данных»

Цель освоения модуля 3 – приобретение слушателями знаний по современным языкам запросов, приобрести умения работы по построению запросов SQL, а также овладеть навыками работы с большими данными с использованием SQL

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 1:

- ПК-1. Применяет стандарты и методики бизнес-моделирования.
- ПК-2. Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.

Планируемые результаты обучения по модулю 1

По итогам освоения модуля слушатели должны:

Знать:

- современные языки запросов;
- источники информации, в том числе информации, необходимой для обеспечения деятельности в предметной области заказчика исследования;
- типы больших данных: метаданные, полуструктурированные, структурированные, неструктурированные;
- виды источников данных: созданные человеком, созданные машинами.

Уметь:

- применять язык SQL для работы по построению запросов, сбора и подготовки данных и визуализации данных;
- собирать, классифицировать, систематизировать и обеспечивать хранение и актуализацию информации бизнес-анализа;
- использовать современные информационные технологии аналитики для информационно-аналитического сопровождения деятельности организации.

Владеть:

- анализом решений с точки зрения достижения целевых показателей решений;
- навыками работы с сырыми данными, использования SQL для интеграции и очистки данных;

– навыками применения технологий машинного обучения для решения анализа данных и принятия решений.

Учебно-тематический план модуля 3 «Языковые средства для аналитики данных»

№ п/п	Наименование дисциплины, модуля, темы	Трудоем- кость		В том числе				Форма контроля
		В зачетных единицах	В часах	Всего	Контактная работа ⁵		Самостоятельная работа*	
					Лекции	Практичес- кие		
1	2	3	4	5	6	7	8	9
	Тема 1. Введение в языки запросов		20	10	4	6	10	
	Тема 2. Построение запросов SQL		15	10	4	6	5	
	Тема 3. SQL в работе с большими данными		17	12	6	6	5	
	Промежуточная аттестация		2	2	-	2		-
	Итого по модулю 3		52	32	14	18	20	2, тест

Содержание модуля 3 «Языковые средства для аналитики данных»

Тема 1. Введение в языки запросов

Понятие «язык запросов». Классификация языков запросов. Особенности языков запросов разных классов. Классификация запросов. Особенности запросов разных классов. Общая характеристика табличных языков запросов. Особенности реализации табличных языков запросов в современных СУБД. Задание простых и сложных запросов. Влияние типа поля на формулирование запроса. Возможности совместной обработки нескольких файлов, связывание файлов. Вычисляемые поля. Упорядочение данных в ответе. Состав и порядок следования полей в ответе. Возможности группировки данных, получение подытогов. Использование агрегатных функций.

Рассматривается практический кейс «Практические примеры баз данных и настройка запросов к ним»

Описание:

Создание базы данных в LibreOffice с помощью PostgreSQL, создание таблиц и заполнение их данными.

Вариант 1 Ателье.

⁵ С применением дистанционных образовательных технологий и электронного обучения

Необходимо хранить информацию о существующем ателье (номер ателье, название, адрес, телефон), видах предоставляемых услуг (код, название, длительность выполнения).

Вариант 2 Частная поликлиника.

Необходимо хранить информацию о врачах (табельный номер, фамилия, специализация, номер кабинета), пациентах (номер страхового полиса, фамилия, адрес, год рождения), а также о талонах на прием с указанием даты, времени приема и суммы страховой выплаты

Вариант 3 Предприятие

Необходимо хранить информацию о сотрудниках предприятия (табельный номер, фамилия, должность, стаж), выполняемых проектах (код проекта, название, дата начала, стоимость), а также об участии сотрудников в конкретных проектах с указанием длительности работы над проектом и оплате.

Вариант 4 Библиотека

Необходимо хранить информацию об абонентах (номер читательского билета, фамилия, адрес, телефон), хранимых книгах (шифр, автор, название, количество экземпляров), а также о выдаче книг на руки с указанием даты выдачи и срока возврата, размере штрафа за утерю книги.

Вариант 5 Склад

Необходимо хранить информацию о существующих складах (номер склада, адрес, телефон, фамилия руководителя склада), товарах (код, название, группа товара, фирма-производитель), а также о наличии товаров на конкретных складах с указанием количества товаров за последние три дня.

Вариант 6 Учебная нагрузка

Необходимо хранить информацию о преподавателях (табельный номер, фамилия, должность, кафедра, стаж), дисциплинах (код, название, направление (гуманитарное, техническое и т.д.)), а также о распределении нагрузки по преподавателям с указанием номера группы студентов, семестра и количества часов.

Вариант 7 Салон красоты

Необходимо хранить информацию о предоставляемых в салоне видах услуг (код, название, цена), клиентах (фамилия, адрес, контактный телефон, год рождения), а также о предварительной записи клиентов на услуги с указанием даты и времени посещения.

Вариант 8 Сеть компьютерных магазинов

Необходимо хранить информацию о существующих магазинах (номер магазина, название, адрес, телефон), комплектующих (модель, название вида,

фирма-производитель, цена), а также о наличии комплектующих в конкретных магазинах с указанием их количества.

Вариант 9 Туристические фирмы города

Необходимо хранить информацию о существующих в городе фирмах (код, название, адрес, фамилия директора), постоянных клиентах (табельный номер, фамилия, адрес, телефон), а также о турах, заказанных клиентами с указанием страны, даты отъезда и стоимости тура/

Вариант 10 Сеть магазинов по продаже бытовой техники

Необходимо хранить информацию о существующих магазинах (номер, адрес, телефон, Ф.И.О. директора, количество сотрудников), о продаваемой технике (код, название, марка, масса, цена), а также о наличии техники в магазинах с указанием количества за последние три дня.

Обращение к базам данных посредством QBE- и SQL- заросов.

Пример практического задания:

Составить запросы с помощью дизайнера.

Вариант 1 Ателье

1. Вывести полную информацию об ателье.
2. Вывести информацию об ателье 1 и 3.
3. Вывести информацию об услугах с длительностью выполнения от 3 до 7.
4. Вывести информацию о ателье отсортировав ее по названию в алфавитном порядке.
5. Найти максимальную длительность услуги.

Тема 2. Построение запросов SQL

Введение в SQL. Стандарты. Диалекты. Составные части языка SQL (DDL, DML, DCL). Расширения языка. Визуальные построители запросов. Структура языка и синтаксис основных операторов SQL. Запросы и подзапросы. Управление выполнением сценария.

Использование SQL для выборки данных из таблиц. Использование булевых и реляционных операторов создания более сложных предикатов. Использование специальных операторов в «условиях». Суммирование данных с помощью функций агрегирования. Форматирование результатов запросов. Использование множества таблиц в одном запросе. Вложенные запросы. Связанные подзапросы. Использование операторов EXIST, ANY, ALL, SOME. Использование предложения UNION. Ввод, удаление, изменение значений полей. Использование подзапросов с командами обновления. Создание таблиц, команда CREATE. Табличные и скалярные функции. Представления. Работа с метаданными на основе системных представлений. Динамическое формирование запросов. Планы выполнения запросов. Корректирующие запросы. Запоминание и корректировка запросов. Запросы на создание таблиц. Табличные языки как «построители» SQL-запросов.

Рассматривается практический кейс «Запросы SQL к базам данных»
Описание. К построенной ранее базе данных построить SQL запросы
Пример задания.

Вариант 1 Ателье

1. Вывести полную информацию из Таблицы Услуги.
2. Вывести информацию название и телефон из Таблицы Ателье
3. Вывести информацию из Таблице Услуги в название поля

Длительность выполнения

добавьте единицы измерения час

4. Вывести первые три записи из Таблицы Ателье

5. Вывести из Таблицы Услуги все данные и добавьте вычисляемое поле

Срочное

выполнение значение, в котором равно Длительность выполнения
деленное на 3.

6. Вывести информацию из Таблицы Услуги Длительность выполнения
больше 10

7. Вывести информацию из Таблицы Услуги Длительность выполнения
больше 10 и

меньше 30

8. Перепишите 7 задание используя инструкцию BETWEEN

9. Вывести информацию об Ателье Золушка, Карина, Мода из Таблицы
Ателье

10. Вывести информацию о Длительности выполнение услуги Подгонка

11. Вывести информацию из об Ателье название которых заканчивается
на букву “а”

Вариант 2 Частная поликлиника

1. Вывести информацию о врачах, у которых указан кабинет.

2. Вывести информацию о пациентах, у которых не указан адрес

3. Вывести список пациентов в алфавитном порядке.

4. Вывести список врачей, отсортировав его по убыванию номеров
кабинетов.

5. Вывести три самых возрастных пациента.

6. Вывести все фамилии пациентов и врачей

7. Выведите информацию о пациенте, дате приема, стоимости приема

8. Выведите информацию о врачах, пациентах, дате приема, стоимости
приема

9. Выведите информацию о врачах, пациентах, дате приема, и стоимости
приема больше

20

Тема 3. SQL в работе с большими данными

Введение в язык SQL-процедур. Параметрические SQL-процедуры.
Системные переменные и способы их вывода. Определение и назначение
триггеров. Примеры триггеров, обеспечивающих целостность и контроль

вводимых данных. Запросы с условиями, которые включают специальные операторы IN, BETWEEN, ALL, LIKE, IS NULL. Запросы с функциями агрегирования: SUM, MAX, MIN и AVG. Использование оператора GROUP BY. Оператор HAVING и особенности его применения с оператором GROUP BY. Использование функций для получения вычисляемых столбцов. Вложенные запросы с оператором EXISTS. Команды DML (INSERT, DELETE и UPDATE). Команда создания таблиц CREATE TABLE.). Контроль вводимых данных: проверка данных на NULL (NOT NULL), проверка на уникальность (UNIQUE), проверка данных по условию (CHECK и CHECK <с условиями>), присвоение значения по умолчанию (DEFAULT). Управление пользователями. Создание резервных копий БД. Миграция данных между различными системами (используя DTS). Экстракция БД в виде SQL-скрипта. Распределение привилегий и защита от несанкционированного доступа: «логин» и пароль» к базе данных, команда GRANT, аргументы ALL и PUBLIC, передача привилегий с использованием GRANT OPTION. Лишение привилегий пользователя (REVOKE). Изучение процедуры инсталляции и настройки сетевого доступа для клиент-серверных СУБД, например, MS SQL Server или Oracle. Организация переноса данных между различными серверами БД с использованием DTS (Data Transformation Services).

Рассматривается практический кейс «Запросы SQL к большим данным»

Пример задания.

Вариант 1 Ателье

1. Создайте представление с полями услуга и стоимость (На основе задание 7 из практики 4)
2. Примените к представлению группировку по услуги, выведите среднюю стоимость услуги
3. Примените фильтр HAVING, выведите услуги, среднюю стоимость больше 100
4. Разработать функцию, возвращающую список ателье, которые выполняют заданную услугу.
5. Разработать хранимую процедуру, которая изменяет длительность услуги заданного пользователем названия услуги, на заданную длительность.
6. Разработать триггер, который позволяет удалять информацию обо всех ателье из таблицы Ателье, кроме ателье с номером 1.

Пример задания

Создать базу данных (БД) на тему «Поставка товаров на склад» - (название БД delivery)

Практика использования функций SELECT DISTINCT, COUNT(), AVG(), SUM(), MIN(), MAX(); предложений GROUP BY и HAVING, представления CREATE VIEW; операторов DML (SELECT, INSERT, UPDATE, DELETE)

Содержание семинаров, практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено занятие семинарского типа	Формы и методы проведения
1	Тема 1. Введение в языки запросов	Решение практических заданий
2	Тема 2. Построение запросов SQL	Решение практических заданий
3	Тема 3. SQL в работе с большими данными	Решение практических заданий

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
1.1.	Тема 1. Введение в языки запросов	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.2.	Тема 2. Построение запросов SQL	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.3.	Тема 3. SQL в работе с большими данными	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации

1. Дайте определение реляционной База Данных? Роль системы управления базами данными (СУБД) в организации.

2. Архитектура современных СУБД. В чем суть логического и инфологического моделирования?

3. Этапы проектирования СУБД. Каковы принципы построения ER-модели?
4. Нормализация БД. В чем особенности первой, второй и третьей нормальной формы?
5. Какие задачи по проектированию БД могут решаться с помощью инструментальных средств ErWin/BrWin.
6. Каково назначение языков DDL и DML. Как создать таблицу БД используя DDL.
7. Какие существуют операторы языка SQL для чтения и записи данных?
8. Какие существуют операторы языка SQL для обновления (редактирования) и удаления данных?
9. Какие существуют операторы языка SQL предоставления прав доступа к отдельным объектам БД различным пользователям?
10. Роль интерфейсов ODBC и ADO в проектах разработки приложений баз данных под Windows?
11. Какие основные функции языка программирования PHP используются для организации доступа к объектам СУБД MySQL Server в WEB-приложениях.
12. Как осуществить доступ к БД из приложений под Windows, используя VBA и ADO.
13. Опишите основные этапы проектирования приложений БД под WEB. Как осуществить передачу данных с WEB-страницы серверу БД?
14. Перечислите функции администратора БД. Как осуществить физический перенос объектов БД с данными между различными серверами (например, в MS SQL Server).
15. Перечислите основные инструментальные средства инструментальные средства СУБД MS SQL Server?
16. Каково предназначение инструментального средства инструментального средства служб Reporting Services?
17. Каково назначение оператора BULK INSERT в Transact SQL?
18. Каково назначение функций – указателей в Transact SQL?
19. Перечислите основные инструментальные средства инструментальные средств СУБД Oracle?
20. Какие функции управления данными могут быть реализованы с помощью Oracle enterprise manager?
21. Из каких ключевых блоков состоит программа PL/SQL?
22. Как осуществить запрос нескольких строк из базы данных, используя курсор PL/SQL?
23. Какие задачи, связанные с проектированием СУБД MySQL Server могут быть выполнены с помощью программы EMS SQL Manager for MySQL?
24. Каковы основные причины внедрения ХД в организации? Чем ХД отличается от транзакционной БД?
25. Опишите основные особенности ROLAP, HOLAP и MOLAP.
26. Опишите схему типа «звезда» и «снежинка».
27. Особенности метода DataVault.

28. Какие типы инфо-кубов поддерживаются в современных ХД.
29. Зачем используются агрегаты в ХД.
30. Расскажите об основных функциональных возможностях языка MDX.
31. Опишите принципы формирования OLAP-отчетности.
32. Опишите процедуру создания аналитического хранилища данных в среде MS SQL Server (SSAS).
33. Что такое ETL-процедура? Опишите возможную схему загрузки данных в ХД из плоских файлов.

Перечень основной и дополнительной учебной литературы

Нормативно-правовые акты:

1. Национальная программа "Цифровая экономика Российской Федерации", утв. распоряжением Правительства Российской Федерации от 28 июля 2021 г. № 1632-р
2. Федеральная проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 N 17).

Основная литература:

1. Сурова Н.Ю. Искусственный интеллект. Монография. Гриф НИИ образования и науки. Гриф МУМЦ "Профессиональный учебник" Гриф МНИЦ Судебной экспертизы и исследований – 2 – е издание М.: Изд-во ЮНИТИ, 2021, 360 с.
2. Аллен, Г. Тейлор SQL для чайников / Аллен Г. Тейлор. - М.: Диалектика, Вильямс, 2021. - 416 с
3. Бен, Форта SQL за 10 минут / Форта Бен. - М.: Диалектика / Вильямс, 2021.
4. Гудсон, Джон Практическое руководство по доступу к данным (+ DVD-ROM) / Джон Гудсон, Роб Стюард. - М.: БХВ-Петербург, 2022. - 304 с.
5. Дейт, К. Дж. SQL и реляционная теория. Как грамотно писать код на SQL / К.Дж. Дейт. - М.: Символ-плюс, 2021.
6. Джеймс, Р. Грофф SQL. Полное руководство / Джеймс Р. Грофф, Пол Н. Вайнберг, Эндрю Дж. Оппель. - М.: Вильямс, 1. - 960 с.
7. Дунаев, В. В. Базы данных. Язык SQL для студента / В.В. Дунаев. - М.: БХВ-Петербург, 2021. - 288 с.
8. Дунаев, В.В. Базы данных. Язык SQL для студента / В.В. Дунаев. - М.: БХВ-Петербург, 2021.
9. Карвин, Билл Программирование баз данных SQL. Типичные ошибки и их устранение / Билл Карвин. - М.: Рид Групп, 2022. - 336 с
10. Кригель, А. SQL. Библия пользователя / А. Кригель. - М.: Диалектика / Вильямс, 2021.

11. Майкл, Дж. Хернандес SQL - запросы для простых смертных. Практическое руководство по манипулированию данными в SQL / Майкл Дж. Хернандес, Джон Л. Вьескас. - М.: ЛОРИ, 2022. - 458 с
12. Майкл, Дж. Хернандес SQL-запросы для простых смертных. Практическое руководство по манипулированию данными в SQL / Майкл Дж. Хернандес, Джон Л. Вьескас. - М.: ЛОРИ, 1. - 480 с
13. Майкл, Мак-Локлин Oracle Database 11g. Программирование на языке PL/SQL / Мак-Локлин Майкл. - М.: ДМК Пресс, 2021.
14. Маркин, А. В. Построение запросов и программирование на SQL. Учебное пособие / А.В. Маркин. - М.: Диалог-Мифи, 1. - 384 с.
15. Мартишин, С. А. Проектирование и реализация баз данных в СУБД MySQL с использованием MySQL Workbench. Учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. - М.: Форум, Инфра-М, 2021. - 160 с.
16. Молинаро, Энтони SQL. Сборник рецептов / Энтони Молинаро. - М.: Символ-плюс, 2021.
17. Эйри, Джоунс Функции SQL. Справочник программиста / Джоунс Эйри. - М.: Диалектика / Вильямс, 2022.

Дополнительная литература:

1. Прайс, Джейсон Oracle Database 11g: SQL. Операторы SQL и программы PL/SQL / Джейсон Прайс. - М.: ЛОРИ, 2022. - 660 с.
2. Пржиялковский, В. В. Введение в Oracle SQL / В.В. Пржиялковский. - М.: Бином. Лаборатория знаний, Интернет-университет информационных технологий, 2021. - 320 с.
3. Прибыл, Билл Oracle PL/SQL. Для профессионалов / Билл Прибыл. - М.: Питер, 1. - 0 с.
4. Проектирование и реализация баз данных Microsoft SQL Server 2000. Учебный курс MCSE (+ CD-ROM). - М.: Русская Редакция, 2021. - 664 с.
5. Рон, Хардман Oracle Database PL/SQL. Рекомендации эксперта / Хардман Рон. - М.: ЛОРИ, 1. - 0 с
6. Саураб, Гупта Oracle PL/SQL. Руководство для разработчиков / Гупта Саураб. - М.: ЛОРИ, 1.
7. Селко, Джо SQL для профессионалов. Программирование / Джо Селко. - М.: ЛОРИ, 2021. - 464 с.
8. Тейлор, Аллен SQL для чайников / Аллен Тейлор. - М.: Вильямс, 1. - 416 с.
9. Форта, Бен Освой самостоятельно SQL за 10 минут / Бен Форта. - М.: Вильямс, 2021. - 288 с.
10. SQL - язык реляционных баз данных: Учебное пособие / Н.Н. Карзаева. - М.: Финансы и статистика, 2022. - 240 с.

Интернет-ресурсы

11. <http://h2o-release.s3.amazonaws.com/h2o/rel-zermelo/4/index.html> - страница загрузки платформы H2O
12. <https://cloud.google.com/bigquery/docs/sandbox> - страница BigQuery sandbox
13. [http://www.tadviser.ru/index.php/Статья:Большие_данные_\(Big_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)) Статья «Большие данные (Big Data)» [Электронный ресурс]
14. <https://biconsult.ru/products/uchebное-posobie-po-tableau> (дата обращения: 22.03.2021). Учебник по Tableau - система бизнес-анализа (BI).
15. <https://biconsult.ru/products/uchebное-posobie-po-tableau> (дата обращения: 22.03.2021). Учебник по Tableau - система бизнес-анализа (BI).
16. <https://cloud.yandex.ru/services/datalens> – сервис визуализации и анализа данных Яндекс
17. <https://community.cloud.databricks.com/login.html> - страница регистрации Databricks Community Edition
18. <https://rapidminer.com/get-started/> - страница загрузки платформы RapidMiner
19. <https://loginom.ru/download> - страница загрузки платформы Loginom компании BaseGroup Labs
20. <https://powerbi.microsoft.com/ru-ru/desktop/> - страница загрузки платформы Power BI Desktop компании Microsoft
21. <https://public.tableau.com/en-us/s/download> - страница загрузки платформы Tableau Public
22. <https://urait.ru/bcode/413823> (дата обращения: 10.04.2022). Зараменских Е. П. Основы бизнес-информатики: учебник и практикум для бакалавриата и магистратуры / Е. П. Зараменских. – Москва: Издательство Юрайт, 2022. — 407 с. — (Высшее образование). — ISBN 978-5-9916-8210-7. — Текст: электронный // Образовательная платформа Юрайт [сайт]
23. <https://www.knime.com/knime-analytics-platform> - страница загрузки платформы Knime Analytics Platform

Описание системы оценки качества освоения дисциплины

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в ходе проведения практических занятий.

Пример практических заданий:

Задание 1.

Дана база данных, состоящая из четырех таблиц:

Клиент (Код_клиента, Фамилия, Имя, Отчество, Город);

Проживает (Код_клиента Номер, Дата_прибытия, Дата_убытия); Номер (Номер, Число мест, Этаж);

Тип номера (Число мест, Цена);

1. Выдать список номеров пятого этажа.
2. Список номеров, в которых останавливался Суворов из Владимира.
3. Каково среднее количество мест в номерах гостиницы?
4. Количество клиентов из каждого города, приезжающих в гостиницу.
5. Выдать список двухместных номеров.
6. Фамилии клиентов, проживающих в настоящее время в гостинице.
7. Сколько всего мест в гостинице.
8. Сколько номеров имеется на каждом этаже гостиницы?

Задание 2.

1. Дана база данных, состоящая из четырех таблиц:

Клиент (Код_клиента, Фамилия, Имя, Отчество, Город);

Проживает (Код_клиента, Номер, Дата_прибытия, Дата_убытия);

Номер (Номер, Число мест, Этаж);

Тип номера (Число мест, Цена)

Напишите запрос с подзапросом, который перечислит фамилии клиентов, которые в текущем году останавливались в одноместных номерах.

Задание 3.

Даны три таблицы ПРОДАВЕЦ (код продавца, имя, город, комиссионные) ЗАКАЗЧИК (код покупателя, ФИО, рейтинг, город, код_продавца); ПОКУПКА (номер, сумма, дата, код_продавца, код_покупателя). Напишите представление, которое выводит рейтинг и имя каждого заказчика в Москве, делавшего покупки во Владимире.

Задание 4.

Дана база данных, состоящая из четырех таблиц: Продавцы (Код_продавца, ФИО_продавца, город_продавца, комиссионные_продавца, руководитель, план продаж); Заказчики (Код_заказчика, ФИО_заказчика, город_заказчика, рейтинг_заказчика, сумма_кредита); Более высокий рейтинг указывают на большее предпочтение. Заказы (номер_заказа, сумма_заказа, дата_заказа, код_продавца, код_заказчика) Товар (Код_товара, Наименование_товара, цена, количество)

1. Напишите триггер, который при добавлении или обновлении таблицы «Заказчики» проверял условие: значение рейтинга должно быть положительным и не превосходить 500.

2. Напишите триггер для увеличения на 5% от суммы покупки комиссионных продавца, оформившего покупку.

3. Напишите триггер, который при добавлении или обновлении заказа в кредит сравнивал сумму заказа с максимально допустимым кредитом для данного заказчика.

4. Напишите триггер, который запрещает покупку, если она оформляется после 25 числа каждого месяца.

5. Напишите хранимую процедуру, которая увеличивает комиссионные у тех продавцов, сумма продаж которых превышает среднюю сумму продаж всей компании.

6. Напишите хранимую процедуру с параметрами показывающую фамилии и сумму продаж продавцов, работающих под руководством указанного начальника.

7. Напишите хранимую процедуру, которая при выполнении заказа автоматически изменяла количество товара в таблице «Товар»

8. Напишите хранимую процедуру с параметрами, которая выводит рейтинг и имя каждого заказчика в Москве, делавшего покупки во Владимире.

Промежуточный контроль: тест. Слушателю предлагается решить тестовые задания.

Пример Тестовых заданий

1. Какая SQL команда используется для обновления данных в базе?

MODIFY

SAVE

SAVE AS

UPDATE

2. Какая SQL команда используется для удаления данных из базы?

REMOVE

EXTRACT

COLLAPSE

DELETE

3. Какая SQL команда используется для вставки данных в базу?

ADD RECORD

INSERT NEW

INSERT INTO

ADD NEW

4. Как выбрать столбец с названием Цена из таблицы Товар?

SELECT Товар.Цена

SELECT Цена FROM Товар

EXTRACT Цена FROM Товар

REMOVE Товар FROM Цена

5. Как выбрать все поля из таблицы Реализация?

SELECT Реализация

SELECT *. Реализация

SELECT [all] FROM Реализация

SELECT * FROM Реализация

6. Как выбрать все записи из таблицы Сотрудники, где значение поля Должность – преподаватель?

SELECT [all] FROM Сотрудники WHERE Должность LIKE 'преподаватель'

SELECT * FROM Сотрудники WHERE Должность <> 'преподаватель'

```
SELECT * FROM Сотрудники WHERE Должность ='преподаватель'  
SELECT [all] FROM Сотрудники WHERE Должность ='преподаватель'
```

7. Как выбрать все записи из таблицы Товар, где значение поля Название начинается с буквы "с"?

```
SELECT * FROM Товар WHERE Товар ='с'  
SELECT * FROM Товар WHERE Товар LIKE 'с%'  
SELECT * FROM Товар WHERE Товар LIKE '%с'  
SELECT * FROM Товар WHERE Товар ='%с%'
```

8. Оператор OR срабатывает, когда ХОТЬ ОДНО условие истинно. AND когда ВСЕ условия истинны.

Ложь
Истина

9. Как выбрать все записи из таблицы Сотрудники, где значение поля Имя – Петр и Фамилия – Иванов?

```
SELECT Имя ='Петр', Фамилия='Иванов' FROM Сотрудники  
SELECT * FROM Сотрудники WHERE Имя  
<>'Петр' AND Фамилия<>'Иванов'  
SELECT * FROM Сотрудники WHERE Имя  
='Петр' AND Фамилия='Иванов'
```

10. Как выбрать все записи из таблицы Сотрудники, где значение поля Фамилия в алфавитном порядке находится между значениями "Иванов" и "Репин"?

```
SELECT * FROM Сотрудники WHERE Фамилия>'Иванов' AND Фамилия  
<'Репин'  
SELECT Фамилия>'Иванов' AND Фамилия <'Репин' FROM Сотрудники  
SELECT * FROM Сотрудники WHERE Фамилия BETWEEN 'Иванов'  
AND 'Репин'
```

11. Какое SQL выражение используется для извлечения различных значений?

```
SELECT UNIQUE  
SELECT DIFFERENT  
SELECT DISTINCT  
SELECT *
```

12. Какая SQL команда используется для упорядочивания результатов?

```
SORT BY  
ORDER  
ORDER BY  
SORT
```

13. Как выбрать все записи из таблицы Товар, упорядоченных по полю Название в обратном порядке?

```
SELECT * FROM Товар SORT BY 'Название' DESC  
SELECT * FROM Товар ORDER BY Название DESC  
SELECT * FROM Товар SORT 'Название' DESC  
SELECT * FROM Товар ORDER Название DESC
```

14. Как вывести количество записей, хранящихся в таблице Товар?

SELECT COUNT() FROM Товар
SELECT COLUMNS() FROM Товар
SELECT COLUMNS(*) FROM Товар
SELECT COUNT(*) FROM Товар

15. Какая инструкция языка SQL используется для создания запроса на выборку?

SELECT
SELECT ... INTO
INSERT ... INTO
ORDER BY DELETE
UPDATE
TRANSFORM

16. Инструкция языка SQL CREATE TABLE используется для:

создания представления
создания индекса
обновления таблицы
сохранения результатов выборки в новой таблице
создания таблицы
удаления таблицы

17. Какая инструкция языка SQL используется для создания запроса на удаление

SELECT
SELECT ... INTO
INSERT ... INTO
ORDER BY
DELETE
UPDATE
TRANSFORM

18. Инструкция языка SQL DROP TABLE используется для:

создания представления
изменения индекса
обновления таблицы
изменения таблицы
создания таблицы
удаления таблицы

19. Для группировки данных потребуется предложение:

ORDER BY
GROUP BY
HAVING

не одно из вышеперечисленных

20. Чтобы исключить дублирующиеся значения из результата выборки, нужно использовать:

SELECT ALL
SELECT DISTINCT
SELECT DISTINCT ROW

SELECT TOP 10% SELECT
NOT DUPLICATE

Порядок проведения: тестирование проводится с личного компьютера, 10 тестовых вопросов по всем темам дисциплины, 60 мин, количество попыток – 3.

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Промежуточная аттестация проводится на основании балльно-рейтинговой системы. В процессе обучения учебные действия слушателя оцениваются по шкале:

Вид учебного действия	Максимальное количество баллов
Академическая активность (посещаемость, активность на семинарах и т.д.)	15
выполнение всех практических заданий	20
текущая аттестация (по всем темам)	25
промежуточный контроль по дисциплине	30
<i>ВСЕГО</i>	<i>100</i>

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Рабочая программа модуля 4

«Использование языка Python для анализа данных»

Цель освоения модуля 1 – приобретение слушателями знаний по конструкциям языка Python, приобрести умения работы со Python и SQL, а также овладеть навыками работы с Python для анализа данных.

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 1:

- ПК-1. Применяет стандарты и методики бизнес-моделирования.
- ПК-2. Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.

Планируемые результаты обучения по модулю 1

По итогам освоения модуля слушатели должны:

Знать:

- типы данных и базовые алгоритмические конструкции в Python,
- средства языка для работы с массивами и строками, ввода/вывода данных через консоль и текстовые файлы, базовые понятия ООП и принципы их использования,
- способы реализации механизмов ООП в Python,
- принципы создания приложений с графическим интерфейсом, назначение и возможности библиотеки Pygame,
- основные понятия и базовые алгоритмы в области анализа данных и машинного обучения,
- назначение и возможности библиотек Pandas, SEABORN, PlotLy, NumPy и Matplotlib;

Уметь:

- использовать интегрированную среду разработки для написания и отладки программного кода на языке Python,
- описывать классы и иерархию связей между ними,
- реализовывать механизмы перегрузки операторов,
- документировать классы,
- устанавливать и подключать к проекту библиотеки,
- использовать структуры данных и методы библиотек Pygame, Pandas, SEABORN, PlotLy, NumPy и Matplotlib;

Владеть:

- навыками работы в интегрированной среде разработки для Python, разработки программ с использованием механизмов ООП,
- навыками разработки графического интерфейса пользователя, навыками решения простейших задач классификации и анализа данных на языке Python.

Учебно-тематический план модуля 4
«Использование языка Python для анализа данных»

№ № п/п	Наименование дисциплины, модуля, темы	Трудоем- кость		В том числе				Форма контроля
		В зачетных единицах	В часах	Всего	Контактная работа ^б из них		Самостоятельная работа*	
					Лекции	Практичес- кие		
1	2	3	4	5	6	7	8	9
	Тема 1. Основные конструкции языка Python		14	10	4	6	4	
	Тема 2. Взаимодействие языка Python и SQL		10	6	2	4	4	
	Тема 3. Возможности Python для анализа данных		10	6	2	4	4	
	Промежуточная аттестация		2	2	-	2		-
	Итого по модулю 1		34	22	8	14	12	2, тест

Содержание модуля 1
«Использование языка Python для анализа данных»

Тема 1. Основные конструкции языка Python

Установка языка Python и необходимых библиотек. Введение в алгоритмизацию и программирование на высокоуровневом языке. Язык Python: особенности, преимущества, «дзен Python». Основы программирования: переменные, присваивание, типы данных, арифметические операции, выражения, функции, ветвления и циклы, объекты и методы, подключение модулей. Основные конструкции языка Python. Введение в Python. Синтаксис языка. Базовые типы данных: числа, строки, списки, кортежи, словари, множества. Функции. Итераторы и генераторы. Классы и объекты. Декораторы. Ввод-вывод. Обработка исключений. IPython, Jupyter Notebook. Подключение библиотек, создание собственных модулей. Элементы функционального программирования (lambda, map, zip, reduce, filter). Чтение и запись данных в текстовом формате.

Рассматривается практичекый кейс

^б С применением дистанционных образовательных технологий и электронного обучения

Дата-сеты: Титаник.csv и Качество воздуха.csv

Тема 2. Взаимодействие языка Python и SQL

Использование Python и SQL вместе. Настройка среды Python и сервер MySQL. Подключение к серверу MySQL на Python. Создание новой базы данных. Создание таблиц и связей между таблицами. Заполнение таблиц данными. Чтение данных. Обновление записей. Удаление записей. Создание записей из списков Python. Создание повторно используемых функций, которые будут делать все это за нас в будущем.

Тема 3. Возможности Python для анализа данных

Основы NumPy: многомерные массивы и векторные вычисления. Индексирование и вырезание. Универсальные функции: быстрые поэлементные операции над массивами. Обработка данных с применением массивов. Методы булевых массивов. Сортировка. Устранение дубликатов и другие теоретико-множественные операции. Файловый ввод-вывод массивов. Линейная алгебра. Генерация случайных чисел. Визуализация данных в Python. Обзор библиотек: matplotlib, pandas, seaborn, plotly, bokeh. Базовые типы визуализаций: графики, столбчатые диаграммы, гистограммы, точечные диаграммы (scatter plots), ящики с усами. Комбинирование различных графических элементов. Построение интерактивных диаграмм с помощью plotly и bokeh. Обработка данных с применением массивов. Методы булевых массивов. Сортировка. Устранение дубликатов и другие теоретико-множественные операции. Файловый ввод-вывод массивов. Решение задач линейной алгебры. Генерация случайных чисел. Визуализация данных с помощью библиотек matplotlib, seaborn. Построение графиков, столбчатых диаграмм, гистограмм, точечных диаграмм (scatter plots), ящиков с усами. Комбинирование различных графических элементов. Построение интерактивных диаграмм с помощью plotly и bokeh.

Решение практического кейса «Использование библиотек Python для анализа данных»

Примеры задач:

Загрузите данные из дата-сета Titanic.csv с помощью pandas. Выведите первые 15 строк таблицы. Соберите новый набор данных, который будет содержать информацию только о пассажирах, севших на корабль в порту Q (Embarked = Q).

Загрузите данные из дата-сета Titanic.csv с помощью pandas. Выведите описательную статистику по датафрейму. Найдите максимальные (max) значения возраста (Age) в зависимости от пола (Sex).

Загрузите данные из дата-сета Titanic.csv с помощью pandas. Определите число пропущенных значений в каждом столбце. Замените все пропущенные значения в столбце возраст (Age) на среднее значение возраста среди мужчин. Проверьте, остались ли пустые значения в этом поле.

Загрузите данные из дата-сета Titanic.csv с помощью pandas. Определите по столбцу Survived (выжил или нет), у кого больше вероятность выжить: у людей старше 30 лет или младше.

Содержание семинаров, практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено занятие семинарского типа	Формы и методы проведения
1	Тема 1. Основные конструкции языка Python	Решение практических заданий
2	Тема 2. Взаимодействие языка Python и SQL	Решение практических заданий
3	Тема 3. Возможности Python для анализа данных	Решение практических заданий

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
1.1.	Тема 1. Основные конструкции языка Python	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.2.	Тема 2. Взаимодействие языка Python и SQL	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
1.3.	Тема 3. Возможности Python для анализа данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации

1. Язык Python и особенности его стиля программирования. Интерактивный режим Python.
2. Синтаксис и управляющие конструкции языка Python. Переменные, значения и их типы. Типы данных в Python.
3. Встроенные операции и функции. Основные алгоритмические конструкции.
4. Условный оператор. Множественное ветвление.
5. Циклы и счетчики.
6. Определение функций. Параметры и аргументы. Вызовы функций. Оператор возврата. Конструкции `*args`, `**kwargs`.
7. Списки, кортежи и словари.
8. Операторы общие для всех типов последовательностей.
9. Специальные операторы и функции для работы со списками. Срезы.
10. Использование библиотеки `pandas` и ее структур данных для загрузки, обработки и хранения данных.
11. Инструменты визуализации данных для Python.
12. Библиотеки и функции Python для статистической обработки и анализа данных.

Решение практических задач.

13. Библиотека `pandas`. Введение в структуры данных `pandas`
14. Объекты `Dataframe` и `Series`.
15. Решение основных задач машинного обучения с применением классических алгоритмов, ансамблей и нейронных сетей в Python. Использование уже обученных нейросетевых моделей для генерации изображений и текстов.
16. Библиотеки и функции Python для работы с временными рядами.
17. Решение оптимизационных задач в Python.

Перечень основной и дополнительной учебной литературы

Нормативно-правовые акты:

1. Национальная программа "Цифровая экономика Российской Федерации", утв. распоряжением Правительства Российской Федерации от 28 июля 2017 г. No 1632-р

2. Федеральной проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 N 17).

Основная литература:

1. Сурова Н.Ю. Искусственный интеллект. Монография. Гриф НИИ образования и науки. Гриф МУМЦ "Профессиональный учебник" Гриф МНИЦ Судебной экспертизы и исследований – 2 – е издание М.: Изд-во ЮНИТИ, 2021, 360 с.
2. Пол Бэрри. Изучаем программирование на Python. – Москва: Издательство «Бомбора», 2021. – 611 с.
3. Майкл Доусон. Програмируем на Python. – Санкт-Петербург: Издательство «Питер, 2020. – 416 с.
4. Марк Лутц. Изучаем Python. – Санкт-Петербург: Издательство «Символ Плюс», 2022. – 830 с.
5. Библиотека PyGame [Электронный ресурс]. – URL: <https://webstart.top/ru/progru/pythonru/pygameru>
6. PyGame и разработка игр [Электронный ресурс]. – URL: <https://younglinux.info/pygame/pygame>
7. Эрик Мэтиз Изучаем Python. Программирование игр, визуализация данных, веб-приложения. – СПб: Питер, 2021. – 496 с.
8. Лутц М. Изучаем Python, 4-е издание. – Пер. с англ. – СПб.: Символ-Плюс, 2022. – 1280 с.
9. Златопольский Д.М. Основы программирования на языке Python. – М.: ДМК Пресс, 2021. – 284 с.
10. Лутц М. Программирование на Python, том II, 4-е издание. – Пер. с англ. – СПб.: Символ-Плюс, 2022. – 992 с.
11. Гэддис Т. Начинаем программировать на Python. – 4-е изд.: Пер. с англ. – СПб.: БХВ-Петербург, 2020. – 768 с.
12. Лучано Рамальо Python. К вершинам мастерства. – М.: ДМК Пресс, 2021. – 768 с.
13. Свейгарт, Эл. Автоматизация рутинных задач с помощью Python: практическое руководство для начинающих. Пер. с англ. — М.: Вильямс, 2021. – 592 с.
14. Рейтц К., Шлюссер Т. Автостопом по Python. – СПб.: Питер, 2021. – 336 с.: ил. – (Серия «Бестселлеры O'Reilly»).
15. Любанович Билл Простой Python. Современный стиль программирования. – СПб.: Питер, 2021. – 480 с.: – (Серия «Бестселлеры O'Reilly»).
16. Шелудько, В. М. Основы программирования на языке высокого уровня Python: учебное пособие / В. М. Шелудько. – Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2021. – 146 с. – ISBN 978-5-9275-2649-9. – Текст: электронный // Электронно-библиотечная

система IPR BOOKS: [сайт]. – URL: <http://www.iprbookshop.ru/87461.html> (дата обращения: 13.02.2020)

17. Доусон М. Програмуем на Python. – СПб.: Питер, 1. – 416 с.
18. Прохоренок Н.А. Python 3 и PyQt. Разработка приложений. – СПб.: БХВ-Петербург, 2022. – 704 с.
19. Прохоренок Н.А. Самое необходимое. — СПб.: БХВ-Петербург, 2022. — 416 с.
20. Амоа К.А. Разработка программных пакетов на языке Python: учебное пособие / К.А. Амоа, Н.А. Рындин, Ю.С. Скворцов. – Воронеж: Воронежский государственный технический университет, ЭБС АСВ, 2020. – 61 с. // Электронно-библиотечная система IPR BOOKS: [сайт].
21. Маккинли Уэс. Python и анализ данных / УэсМаккинли; пер. А. Слинкина. – 2-е изд. – Саратов: Профобразование, 2020. – 482 с. 3. Сузи Р.А. Язык программирования Python: учебное пособие / Р.А. Сузи. – 3-е изд. – М.: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. – 350 с. // Электронно-библиотечная система IPR BOOKS: [сайт].
22. Федоров Д.Ю. Программирование на языке высокого уровня Python: учебное пособие / Д.Ю. Федоров. – 2-е изд.– М.: Юрайт, 2020. – 161 с.

Дополнительная литература:

23. Рашка С. Python и машинное обучение –Москва: ДМК Пресс, 2021. – 418 с.
24. Коэльо Л.П. Построение систем машинного обучения на языке Python – Москва: ДМК Пресс, 2021. – 302 с.
25. Остроух А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. – Санкт-Петербург: Лань, 2021. – 228 с.
26. Лутц М. Программирование на Python, том I, 4-е издание. – Пер. с англ. – СПб.: Символ-Плюс, 2022. – 992 с.
27. Шелудько, В. М. Язык программирования высокого уровня Python. Функции, структуры данных, дополнительные модули: учебное пособие / В. М. Шелудько. – Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2021. – 107 с. – ISBN 978-5-9275-2648-2. – Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. – URL: <http://www.iprbookshop.ru/87530.html> (дата обращения: 13.02.2020).
28. Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для прикладного бакалавриата / Д. Ю. Федоров. – 2-е изд., перераб. и доп. – Москва: Издательство Юрайт, 2020. – 161 с. – (Бакалавр. Прикладной курс). – ISBN 978-5-534-10971-9. – Текст: электронный // ЭБС Юрайт [сайт]. – URL: <https://urait.ru/bcode/437489>
29. Гуриков С. Р. Основы алгоритмизации и программирования на Python : учеб. пособие / С.Р. Гуриков. М.: ФОРУМ: ИНФРА-М, 2022. 343 с. (Высшее образование: Бакалавриат). - Режим доступа: <http://znanium.com/catalog/product/924699>
30. Саммерфилд М., Python на практике [Электронный ресурс] / Марк Саммерфилд - М.: ДМК Пресс, 1. - 338 с. - ISBN 978-5-97060-095-5

31. Волкова В.М., Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python [Электронный ресурс]: учебное пособие / Волкова В.М. - Новосибирск: Изд-во НГТУ, 2021. - 74 с. - ISBN 978-57782-3183-2 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785778231832.html>

Интернет-ресурсы

32. <https://docs.python.org/3/>
33. <https://pythonworld.ru/>
34. <https://metanit.com/python>
35. <https://www.pygame.org/wiki/>
36. <https://www.pygame.org/docs>
37. <https://www.kaggle.com/datasets?topic=trendingDataset>
38. <https://archive.ics.uci.edu/ml/index.php>
39. <https://data-flair.training/blogs/machine-learning-datasets/>

Описание системы оценки качества освоения дисциплины

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в ходе проведения практических занятий.

Пример практических заданий:

Задание 1.

Вычислить множество значений функции $y = ax^2 + bx + c$ (построить таблицу значений) для аргумента x , изменяющегося от -10 до 10 с шагом 2 , при заданных a, b, c .

Задание 2.

Распечатать последовательность $(a_1, a_2, \dots, a_{20})$, элементы которой вычисляются по формуле $a_n = n/(n+1)$.

Промежуточный контроль: тест. Слушателю предлагается решить тестовые задания.

Пример Тестовых заданий

1. Коэффициент вариации веса чебурека равен 15% , а средний вес в некоторой выборке чебуреков составил 160 г. Чему равна выборочная дисперсия веса чебурека?

- 1) 24
- 2) $0,15$
- 3) 576
- 4) Недостаточно данных для решения

2. Чему равен двусторонний квантиль уровня 0,95 распределения Стьюдента, если объем выборки равен 38?

- 1) 1,69
- 2) 2,02
- 3) 1,68
- 4) 2,03

3. Что показывает правосторонний доверительный интервал для числовой характеристики θ уровня γ ?

- 1) Что с вероятностью γ значение θ попадает в диапазон от α до β
- 2) Что с вероятностью γ значение θ попадает в диапазон от $-\infty$ до β
- 3) Что при уровне надежности γ значение θ не попадает в диапазон от α до β
- 4) Что при уровне надежности γ значение θ не попадает в диапазон от $-\infty$ до β

4. Грибник пошел в лес собирать грибы. Его задача – определить, является ли каждый из встреченных им грибов ядовитым. Он выдвигает нулевую гипотезу о том, что гриб ядовитый, и в зависимости от результата проверки этой гипотезы собирает или оставляет гриб. Что будет являться ошибкой второго рода?

- 1) игнорирование подберезовиков
- 2) сбор поганок
- 3) ошибочное принятие мухоморов за поганки
- 4) ошибочное принятие опят за маслята, а синявок за рыжики

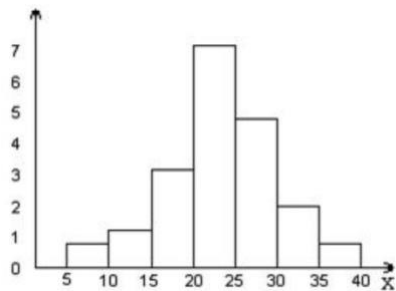
5. Проверяется гипотеза о том, что объем кредитной нагрузки жителя России является нормально распределенной случайной величиной. По результатам анализа выборки из 50 жителей было получено р-значение = 0,08. Какое утверждение является верным?

- 1) на уровне значимости 1% можно утверждать, что объем кредитной нагрузки имеет нормальное распределение
- 2) верны все утверждения
- 3) на уровне значимости 5% недостаточно оснований для отклонения нулевой гипотезы
- 4) на уровне значимости 10% можно утверждать, что объем кредитной нагрузки не имеет нормального распределения

6. Есть ли выбросы в данной выборке? Определить с помощью "ящика с усами" (box plot) 5, -3, -19, 9, 1, 24, -3, 8, -1, 5

- 1) есть аномально малые значения
- 2) есть как аномально большие, так и аномально малые значения
- 3) есть аномально большие значения
- 4) выбросов нет

7. Дана гистограмма, построенная по интервальному ряду некоторой дискретной случайной величины. Если бы эта случайная величина была непрерывной, какое правдоподобное предположение о законе ее распределения можно было бы выдвинуть?



- 1) Случайная величина имеет нормальное распределение
- 2) Случайная величина имеет распределение хи-квадрат
- 3) Случайная величина имеет распределение Фишера
- 4) Случайная величина имеет распределение Пуассона

8. Дана выборка значений трех показателей: [X]: 12, -18, 12, -30, 24, -30, 18. [Y]: 10.3, 16, 12.6, 24.1, 10.9, 24.3, 18.8. [Z]: 70, -30, 60, -10, -20, 20, -40. Построить корреляционную матрицу для показателей [X], [Y], [Z]. Ответ записать в текстовом виде.

X, Y, Z

1,-0.8,0.11

-0.8,1,-0.36

0.11,-0.36,1

9. Выберите задачи машинного обучения, относящиеся к виду «обучение с учителем».

- 1) Задача классификации
- 2) Задача кластеризации
- 3) Задача регрессии
- 4) Задача понижения размерности

10. Выберите библиотеки Python для построения нейронных сетей.

- 1) Pandas
- 2) Requests
- 3) Keras
- 4) TensorFlow

11. Что из приведенного списка не является примером итерируемого объекта в Python?

- 1) диапазон (range)
- 2) кортеж (tuple)
- 3) цикл (for)
- 4) строка (str)

12. Каким будет результат вычисления следующего выражения на языке Python? $(2 * 3 ** 2 + 2) / 1 ** 2$

- 1) 20 (int)
- 2) 20.0 (float)
- 3) 400 (int)
- 4) 400.0 (float)

Порядок проведения: тестирование проводится с личного компьютера, 10 тестовых вопросов по всем темам дисциплины, 60 мин, количество попыток – 3.

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Промежуточная аттестация проводится на основании балльно-рейтинговой системы. В процессе обучения учебные действия слушателя оцениваются по шкале:

Вид учебного действия	Максимальное количество баллов
Академическая активность (посещаемость, активность на семинарах и т.д.)	15
выполнение всех практических заданий	20
текущая аттестация (по всем темам)	25
промежуточный контроль по дисциплине	30
ВСЕГО	100

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Рабочая программа модуля 5

«Технический анализ данных и подготовка данных к визуализации и анализу»

Цель освоения модуля 5 – приобретение слушателями знаний о задачах и методах технического анализа данных и умения подготавливать данные их дальнейшей визуализации и анализу.

Профессиональные компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения модуля 5:

- ПК-1. Применяет стандарты и методики бизнес-моделирования.
- ПК-2. Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.
- ПК-3. Использует большие данные: знает, понимает и обладает навыками использования разновидностей анализа данных; приемами анализа данных; проверкой гипотез о зависимости данных; методами сбора данных; с графическими данными; с текстовыми данными; с аудио данными; с видео данными и т.д.

В рамках данного модуля приобретаются компетенции универсальной модели компетенций в сфере искусственного интеллекта, разработанной РЭУ им. Г.В. Плеханова в рамках результата Федерального проекта «Искусственный интеллект»:

- ПК-7.п. - Способен выполнять анализ больших данных;
- ПК-7.р. - Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта;

Планируемые результаты обучения по модулю 1

По итогам освоения модуля слушатели должны:

Знать:

- теоретические основы анализа данных, технология OLAP и Data Mining;
- методику KDD (Knowledge Discovery in Databases), используемую при разработке моделей и решении стандартных задач Data Mining.
- правила и методы проектирования моделей данных, внедрения аналитических отчетов, информационных панелей мониторинга в компоненты ИТ инфраструктуры организации;
- алгоритмические и программные методы проведения и интеллектуального анализа данных для принятия взвешенных решений;

Уметь:

- проводить технический анализ данных и готовить данные к визуализации и анализу»;

- решать задачи профессиональной деятельности, используя OLAP технологии аналитики;
- использовать современные информационные технологии аналитики для информационно-аналитического сопровождения деятельности;
- работать с многомерными моделями

Владеть:

- инструментами и технологиями выявления, сбора и подготовки к анализу информации бизнес-анализа для формирования возможных решений;

Учебно-тематический модуля 5

«Подготовка данных к визуализации и анализу»

№ п/п	Наименование дисциплины, модуля, темы	Трудоемкость		В том числе				Форма контроля
		В зачетных единицах	В часах	Контактная работа ⁷			Самостоятельная работа*	
				Всего	из них			
					Лекции	Практические		
1	2	3	4	5	6	7	8	9
	Тема 1. Технический анализ данных и подготовка данных к визуализации и анализу. Обзор систем визуального анализа данных		28	18	4	14	10	
	Тема 2. Визуализация анализа данных		30	20	6	14	10	
	Промежуточная аттестация		2	2	-	2		-
	Итого по модулю 1		58	38	10	28	20	2, тест

Содержание модуля 5

«Подготовка данных к визуализации и анализу»

Тема 1. Технический анализ данных и подготовка данных к визуализации и анализу. Обзор систем визуального анализа данных

Технический анализ данных и подготовка данных к визуализации и анализу. Визуализация данных, цели и задачи. Обзор Business Intelligence как процесс анализа информации, выработки интуиции и понимания для улучшенного и неформального принятия решений бизнес-пользователями, а также инструменты для извлечения из данных значимой для бизнеса информации. Обзор принципов BI как совокупность технологий, программного обеспечения и практик, направленных на достижение целей

⁷ С применением дистанционных образовательных технологий и электронного обучения

бизнеса путём наилучшего использования имеющихся данных. Основные принципы проектирования моделей данных, внедрения аналитических отчетов, дашбордов, информационных панелей мониторинга в компоненты ИТ инфраструктуры организации для проведения мониторинга и контроллинга стратегических целей и поддержку внутренних процессов.

Тема 2. OLAP-анализ

Оперативная аналитическая обработка данных (OLAP, OnLine Analytical Processing), анализ данных в различных срезах «slicing and dicing». Интерактивная визуализация как максимально наглядное представление данных с использованием различных интерактивных изображений, схем и графиков. Предикативное моделирование и data mining (интеллектуальный анализ данных) как инструмент для классифицирования данных, формирования собственных номинальных и количественных шкал. Графические средства – отчеты, графики, диаграммы, настраиваемые при помощи параметров.

Содержание семинаров, практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено занятие семинарского типа	Формы и методы проведения
1	Технический анализ данных и подготовка данных к визуализации и анализу. Обзор систем визуального анализа данных	Решение практических заданий
2	OLAP-анализ	Решение практических заданий

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ темы	Наименование (содержание) темы, по которой предусмотрена самостоятельная работа	Формы и методы проведения
1.1.	Технический анализ данных и подготовка данных к визуализации и анализу. Обзор систем визуального анализа данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

1.2.	OLAP-анализ	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
------	-------------	---

Рекомендуемый перечень вопросов для отработки в часы самостоятельной работы, подготовки к промежуточной аттестации

1. Цели и задачи визуализации данных.
2. Основные задачи OLAP-анализа.
3. OLTP-анализ.
4. Системы поддержки и принятия решений и управленческие информационные системы.
5. Задачи Data Mining. Стадии Data Mining. Виды и методы анализа данных.
6. Сферы применения технологии интеллектуального анализа данных.
7. Одномерная визуализация данных.
8. Двумерная визуализация данных.
9. Многомерная визуализация данных
10. Многомерные, реляционные, гибридные и виртуальные хранилища данных;
11. Методология интеллектуального анализа данных.
12. Визуализаторы для интерпретации результатов OLAP-анализа.
13. Требования к средствам оперативной аналитической обработки.
14. Концептуальное многомерное представление.
15. Способы описания данных.
16. Интеллектуальный анализ данных.
17. Критерии оценки существующих продуктов.
18. Визуализаторы общего назначения, для оценки качества моделей.
19. Программное обеспечение в области анализа данных.
20. Способы визуализации данных.
21. Характеристики средств визуализации данных
22. Интеграция OLAP и ИАД.
23. Классификация продуктов OLAP по способу представления.
24. Классификация продуктов OLAP по способу представления данных
25. Древовидные визуализаторы.
26. ROC-кривые.
27. OLAP-анализ.
28. Двенадцать правил Кодда.
29. Дополнительные правила Кодда.
30. Тест FASMI.

Перечень основной и дополнительной учебной литературы

Нормативно-правовые акты:

Национальная программа "Цифровая экономика Российской Федерации", утв. распоряжением Правительства Российской Федерации от 28 июля 2017 г. No 1632-р

Федеральной проект «Искусственный интеллект» (паспорт) (утв. президиумом Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 N 17).

Основная литература:

1. Сурова Н.Ю. Искусственный интеллект. Монография. Гриф НИИ образования и науки. Гриф МУМЦ "Профессиональный учебник" Гриф МНИЦ Судебной экспертизы и исследований – 2 – е издание М.: Изд-во ЮНИТИ, 2021, 360 с.

2. Боровская Е.В. Основы искусственного интеллекта [Электронный ресурс] / Е.В. Боровская, Н.А. Давыдова. – Москва: Лаборатория знаний (ранее БИНОМ. Лаборатория знаний), 2021. – 127 с. – ЭБС Лань

3. Клетте, Р. Компьютерное зрение. Теория и алгоритмы: учебник [Электронный ресурс] / Р. Клетте ; перевод с английского А. А. Слинкина. – Москва : ДМК Пресс, 2020. – 506 с. – ЭБС Лань

4. Паттерсон, Д. Глубокое обучение с точки зрения практика / Д. Паттерсон, А. Гибсон. — Москва : ДМК Пресс, 2022. — 418 с. — ISBN 978-5-97060-481-6. <https://e.lanbook.com/book/116122>

5. Антонио, Д. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / Д. Антонио, П. Суджит ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2022. — 294 с. — ISBN 978-5-97060-573-8. <https://e.lanbook.com/book/111438>

6. Коэльо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Коэльо, В. Ричарт ; перевод с английского А. А. Слинкин. — 2-е изд. — Москва : ДМК Пресс, 2021. — 302 с. — ISBN 978-5-97060-330-7. <https://e.lanbook.com/book/82818>

7. Джонс, М. Т. Программирование искусственного интеллекта в приложениях / М. Т. Джонс. — Москва : ДМК Пресс, 2022. — 312 с. — ISBN 978-5-94074-746-8. <https://e.lanbook.com/book/1244>

8. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва: ДМК Пресс, 2022. — 652 с. — ISBN 978-5-97060-618-6. <https://e.lanbook.com/book/107901>

Дополнительная литература:

9. Алексеев Д.С., Щекочихин О.В. Технологии интеллектуального анализа данных. Учебное пособие. Кострома, 2020. 140 с. <https://elibrary.ru/item.asp?id=43946965>
10. Жаров А.Н., Минеичева И.Г. Анализ данных. Ярославль, 2020. 148 с. <https://elibrary.ru/item.asp?id=43846458>
11. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с. <https://urait.ru/bcode/432851>
12. Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск : УлГТУ, 2021. — 290 с. — ISBN 978-5-9795-1712-4. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053>
13. Маккинни, У. Python и анализ данных / У. Маккинни; перевод с английского А. А. Слинкина. -- 2-ое изд., испр. и доп. -- Москва: ДМК Пресс, 2020. -- 540 с. -- ISBN 978-5- 97060-590-5. -- Текст: электронный // Лань: электронно- библиотечная система. -- URL: <https://e.lanbook.com/book/131721>
14. Полупанов, Д. В. Программирование в Python 3: учебное пособие / Д. В. Полупанов, С. Р. Абдюшева, А. М. Ефимов. -- Уфа: БашГУ, 2020. -- 164 с. -- ISBN 978-5-7477-5230-6. -- Текст: электронный // Лань: электронно-библиотечная система. -- URL: <https://e.lanbook.com/book/179915>
15. Бизли, Д. Python. Книга рецептов / Д. Бизли, Б. К. Джонс ; перевод с английского -- Текст :электронный // Лань : электронно-библиотечная система. -- URL: <https://e.lanbook.com/book/131723>
16. Груздев, А. В. Изучаем Pandas / А. В. Груздев, М. Хейдт ; перевод с английского А. В. Груздева. -- 2-ое изд., испр. и доп. -- Москва : ДМК Пресс, 2020. -- 700 с. -- ISBN 978-5-97060-670-4. -- Текст : электронный // Лань : электронно- библиотечная система. -- URL: <https://e.lanbook.com/book/131693>

Интернет-ресурсы

<https://metanit.com/python>

<https://www.pygame.org/wiki/>

<https://www.pygame.org/docs>

<https://www.kaggle.com/datasets?topic=trendingDataset>

<https://archive.ics.uci.edu/ml/index.php>

<https://data-flair.training/blogs/machine-learning-datasets/>

Описание системы оценки качества освоения дисциплины

В систему оценки качества освоения программы «Технический аналитик (Technical Analyst Data Science)» входят:

- текущий контроль по каждой теме, где предусмотрены практические занятия;
- промежуточная аттестация по каждому модулю;
- итоговая аттестация.

Контроль результатов освоения дисциплины осуществляется в ходе текущего контроля успеваемости и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в ходе проведения практических занятий.

Пример практических заданий:

1. Составить алгоритм очищения данных.
2. Составить алгоритм визуализации деревьев.
3. Составить алгоритм визуализации ассоциативных правил.
4. Составить алгоритм визуализации кластеризованных данных.
5. Составить алгоритм визуализации карт.
6. Составить структуру корпоративной информационно-аналитической системы.

Промежуточный контроль: тест.

Слушателю предлагается решить тестовые задания.

1) Открытые тесты (необходимо вписать слово в предложение):

- _____ - комплекс методов для визуализации многомерных данных.
- оценка _____ моделей – позволяет оценивать различные характеристики моделей, такие как точность, эффективность, достоверность результатов, интерпретируемость, устойчивость и т.д;
- _____ представляют собой линии, отображающие зависимость между несколькими переменными в некоторой системе координат.

2) Закрытые тесты (необходимо выбрать один):

- Выберете из списка метод визуализации данных упорядоченных в иерархии: Лепестковая диаграмма; OLAP; Деревья.
- Выберете из списка задачу не относящуюся к процессу визуализации данных: · проверка корректности моделей, разработка математической модели.

В качестве примера оценочных средств для текущего контроля успеваемости представлены следующие варианты заданий:

1) Открытые тесты (необходимо вписать слово в предложение):

- _____ – насколько правильно модель может работать со всеми возможными входными данными.
- Оперативная _____ обработка и интеллектуальный анализ данных - две составные части процесса поддержки принятия решений.
- _____ кластеризацию – кластеры большого размера разделяются на несколько кластеров меньшего размера.

2) Закрытые тесты (необходимо выбрать один или несколько вариантов ответа):

- Выберете свойства, относящиеся к двенадцати правилам Кобба
многомерность;
прозрачность;
стационарность;

пакетное извлечение против интеграции.

- Распространённые типы визуализаторов, применяемые для интерпретации результатов ...:

древовидные визуализаторы; ·

сетевые карты кохонена

рекурсивные шаблоны.

Порядок проведения: тестирование проводится с личного компьютера, 10 тестовых вопросов по всем темам дисциплины, 60 мин, количество попыток – 3.

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Промежуточная аттестация проводится на основании балльно-рейтинговой системы. В процессе обучения учебные действия слушателя оцениваются по шкале:

Вид учебного действия	Максимальное количество баллов
Академическая активность (посещаемость, активность на семинарах и т.д.)	15
выполнение всех практических заданий	20
текущая аттестация (по всем темам)	25
промежуточный контроль по дисциплине	30
<i>ВСЕГО</i>	<i>100</i>

Перевод рейтинговых баллов в оценку производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка (зачет)
<65	не зачтено
$65 \leq, 75 <$	зачтено

Организационно-педагогические условия реализации программы повышения квалификации «Технический аналитик (Technical Analyst Data Science)»

1) Применяемые образовательные технологии, формы и методы обучения, в том числе интерактивные

В образовательном процессе используются разнообразные формы работы со слушателями, а именно:

- лекция с мультимедийным сопровождением по наиболее сложным вопросам программы

- информационная лекция по наиболее сложным вопросам программы расширенного формата с использованием современных технических средств обучения;

- творческая мастерская/мастер-класс - позволяет решить задачи: личностного саморазвития; образовательной мотивации: повышения интереса к процессу обучения и активного восприятия учебного материала; функциональной грамотности и креативности: навыков и умений творческого постижения и осмысления нового знания; социальной компетентности: коммуникативных навыков и ответственности за знание.

- кейс-стади (в том числе видео-кейсы)– изучение конкретных ситуаций из практики (casestudy), для выполнения данного вида заданий обучающимся должна быть представлена в письменной форме информация относительно реальной ситуации (профессиональной или жизненной) и поставлены конкретные задачи её изучения проблемы, обучающиеся анализируют различные аспекты проблемы и предлагают выработанные решения;

- практикум – это вид практических занятий тренировочного характера, на котором осуществляется связь изучаемой теории и практики, а материал его часто служит иллюстрацией к лекции;

- разработка презентации – самостоятельная работа с использованием информационных технологий и знаний, полученных на занятиях;

- тестирование - это стандартизированный метод оценки знаний, умений, навыков обучающихся и др.

2) Перечень информационных технологий, используемых при осуществлении образовательного процесса

Осуществляется проведение вебинаров для слушателей в удаленном доступе. Преподавателями используются компьютерные презентации, работа в чате, индивидуальное консультирование слушателей.

Условия для функционирования электронной информационно-образовательной среды

Электронные информационные ресурсы	Вид Занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения, система видеоконференцсвязи	Итоговая аттестация	Компьютер, подключенный к сети Интернет; интернет-браузер

3) Описание материально-технической базы, необходимой для осуществления образовательного процесса

Наименование специализированных учебных помещений	Вид занятий	Наименование оборудования, программного обеспечения
Учебный интерактивный тренажерный класс	Лекции Практические занятия	Мультимедийное оборудование, компьютеры, МФУ. Компьютер, подключенный к сети Интернет, интернет-браузер.

Материально-технические условия соответствуют действующим санитарным и противопожарным правилам и нормам.

Примечание. В случае проведения учебных занятий с применением дистанционных образовательных технологий (ДОТ) у слушателя должен быть персональный компьютер, оснащенный аудиокolonками, с доступом в сеть интернет и установленным видеоплеером, способным воспроизводить видеофайлы.

4) Кадровое обеспечение образовательного процесса

Учебный процесс со слушателями обеспечивают квалифицированные сотрудники Финансового университета, а также приглашенные специалисты и действующие практики других организаций.

Реализация программы обеспечивается преподавательским составом из числа штатных преподавателей и совместителей, удовлетворяющих следующим условиям:

- Наличие высшего профессионального образования, соответствующего профилю программы;
- Значительный опыт практической деятельности в соответствующей сфере.

В преподавательский состав образовательной программы «Технический аналитик (Technical Analyst Data Science)» включены 10 человек – преподавателей и научных сотрудников Финансового университета, а также приглашенных специалистов и действующих практиков других организаций.

5) Выдаваемый документ при успешном освоении программы

Удостоверение о повышении квалификации Финансового университета при Правительстве Российской Федерации.

Описание системы оценки качества освоения программы «Технический аналитик (Technical Analyst Data Science)»

Входная диагностика

Входная диагностика состоит из решения тестового задания из 20 вопросов. Слушатель выбирает утверждения, которые он считает верными. Выбранные утверждения демонстрируют степень его владения компетенциями в области искусственного интеллекта и методов машинного обучения.

Каждому утверждению присваивается балльная оценка (от 0 до 1 балла). Оценивание производится на основании суммы баллов.

При отсутствии ответов оценка 0 (ноль) баллов присваивается по умолчанию.

Шкала оценивания:

От 0 до 10 баллов – входная диагностика считается не пройденной;

От 10 до 15 баллов – входная диагностика считается пройденной.

Таким образом, слушатель считается успешно прошедшим входное тестирование при наборе не менее 50 % правильных ответов.

Тестовое задание. *Правильный ответ выделен полужирным.

1. Укажите верное определение машинного обучения?
 - a) Математическая дисциплина, использующая разделы математической статистики, численных методов оптимизации, теории вероятностей, дискретного анализа, и извлекающая знания из данных
 - b) Технология прикладных систем искусственного интеллекта, в которых параметры моделей вычисляются в ходе интеллектуального процесса обучения**
 - c) Наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ
 - d) Свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека

2. Что из перечисленного не относится к языкам программирования?
 - a) R
 - b) TensorFlow**
 - c) Hackel
 - d) C++

3. Группировки множества объектов на подмножества - это ...
 - a) Кластеризация**
 - b) Корреляция
 - c) Классификация
 - d) Регрессия

4. Какие данные из перечисленных ниже можно отнести к Big Data
- a) Перечень студентов университета
 - b) Каталог товаров гипермаркета
 - c) Логи пользователей портала yandex.ru**
 - d) Ежедневные данные о сейсмической активности в течение года
5. Выбросом в данных является:
- a) Выбранные значения перед построением модели
 - b) Значения равные среднему значению в данных
 - c) Значения, которые выделяются из набора данных**
 - d) Значения в пределах второго и четвертого квартилей
6. Python – это ...
- a) Интерпретируемый язык с динамической типизацией и автоматическим управлением памятью**
 - b) Интерпретируемый язык со статической типизацией и автоматическим управлением памятью
 - c) Компилируемый язык с динамической типизацией и ручным управлением памятью
 - d) Компилируемый язык с динамической типизацией и автоматическим управлением памятью
7. Какое выражение на языке Python является корректным?
- a) `int a = 1`
 - b) `a == 3`
 - c) `var a = 3;`
 - d) `a = 3`**
8. С какого ключевого слова начинается объявление функции в Python?
- a) `func`
 - b) `return`
 - c) `def`**
 - d) `define`
9. Выберите «три кита» ООП
- a) Наследие, Абстракция, Полиморфизм
 - b) Наследование, Инкапсуляция, Полиморфизм**
 - c) Класс, объект, переменная
 - d) `map()`, `filter()`, `reduce()`
10. Семантическая сеть - это ...
- a) Раздел лингвистики, устанавливающий отношения между символами и объектами, которые они обозначают

- b) **Информационная модель предметной области, имеющая вид ориентированного графа**
- c) Математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования сетей нервных клеток живого организма
- d) Система, обеспечивающая обмен данными между вычислительными устройствами и другим оборудованием или программным обеспечением

11. К фреймворкам для машинного обучения относится:

- a) Tensorflow, Keras, Caffe, CUDA, Haskel, ASIC
- b) Tensorflow, Keras, Caffe, PyTorch, CatBoost, Kaggle
- c) **Tensorflow, Keras, Caffe, PyTorch, CatBoost, CUDA**
- d) Tensorflow, Keras, Caffe, PyTorch, CatBoost, OpenVR

12. "1. Каков результат следующего выражения: $8 / 2(2+2) = ?$ "

- a) 1
- b) **16**
- c) 4
- d) 2

13. Выберите лишнее число:

- a) 13
- b) **15**
- c) 17
- d) 19

14. Каков результат следующего выражения: $(3^2 - 2^3 + 2^4 - 4^2) ^ 4 = ?$

- a) 1721891
- b) 0
- c) **1**
- d) 3240

15. Напишите чему равно $4! - 3!$

- a) 0
- b) 1
- c) 20
- d) **18**

16. Молоток и гвозди вместе стоят 1 доллар и 10 центов, известно, что молоток стоит на доллар дороже. Сколько стоит молоток?

- a) 5 центов
- b) **доллар 5 центов**
- c) доллар

- d) 90 центов
17. Чему равна сумма вероятностей противоположных событий?
a) 0
b) 1
c) Находится в диапазоне от 0 до 1
d) Не может быть больше 1
18. Массив (матрица) – это ...
a) Цикл с постусловием
b) Переменная строчного-цифрового формата
c) Структура данных, хранящая набор значений, идентифицируемая по индексу
d) Математический термин, не относящийся к программированию
19. Что такое сортировка пузырьком?
a) Сортировка генерацией всевозможных комбинаций массива
b) Повторяющийся проходов по массиву и сравнение соседних элементов
c) Произвольное перемещение массива с проверкой порядка
d) Поиск наименьшего элемента и помещение его в начало
20. В чём смысл двоичного (бинарного) поиска?
a) Дробление массива на половины
b) Получение чисел из суммы двух предыдущих
c) Повторяющийся проходов по массиву и сравнения соседних элементов
d) Поиск двух элементов из множества

ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

В систему оценки качества освоения программы «Технический аналитик (Technical Analyst Data Science)» входят:

- промежуточная аттестация по каждому модулю;
- итоговая аттестация.

ПРОМЕЖУТОЧНЫЕ АТТЕСТАЦИИ ПО МОДУЛЯМ

Методы текущего контроля:

- *компьютерное тестирование*

Пример тестового вопроса текущего контроля:

1. Аналитик это ...

- a) специалист в области анализа и моделирование
b) специалист в предметной области;

- в) человек, решающий определенные задачи;
- г) человек, который имеет опыт в программировании.

2. Эксперт это ...

- а) специалист в области анализа и моделирование;
- б) специалист в предметной области;
- в) человек, решать определенные задачи;
- г) человек, который имеет опыт в программировании.

3. Задача классификации сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

4. Задача регрессии сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

5. Задача кластеризации заключается в ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

6. Целью поиска ассоциативных правил является ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

7. До предполагаемых моделей относятся такие модели данных:

- а) модели классификации и последовательностей;
- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
- в) классификации, кластеризации, исключений, итоговые и ассоциации;
- г) модели классификации, последовательностей и исключений.

8. В описательных моделей относятся следующие модели данных:

- а) модели классификации и последовательностей;
- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;

- в) классификации, кластеризации, исключений, итоговые и ассоциации;
- г) модели классификации, последовательностей и исключений.

9. Модели классификации описывают ...

- а) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

10. Модели последовательностей описывают ...

- а) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

Порядок проведения: тестирование проводится с личного компьютера, 10 тестовых вопросов по каждой теме, 20 мин, количество попыток не ограничено

Критерии оценивания: для получения промежуточной аттестации по теме необходимо набрать не менее 67% от максимального количества баллов в любой попытке.

- *расчетные задачи*

пример расчетной задачи:

Компанию по прокату автомобилей интересует зависимость между пробегом автомобилей X (тыс. км) и стоимостью технического обслуживания Y тыс.руб. Для выяснения характера связи было отобрано 30 автомобилей и получены оценки:

$\sum Y_i = 30$; $\sum X_i = 90$; $\sum X_i Y_i = 135$; $\sum Y_i^2 = 330$; $\sum X_i^2 = 360$. Оценить тесноту линейной связи между переменными с помощью выборочного коэффициента корреляции.

- *кейс-задание*

ЗАДАНИЕ

Вычислите первые 100 чисел Фибоначчи. (Напишите код.)

```
def fibonacci(n=10, a=0, b=1):
    yield a
    yield b
    n -= 2
    while n > 0:
        c = a + b
        a = b
        b = c
        yield c
        n -= 1

if __name__ == '__main__':
    for n in fibonacci(100):
        print(n)
```

Порядок проведения: вариант кейс-задания выдает преподаватель; защита решения кейс-задания проводится в режиме онлайн.

Критерий оценивания: Оценку выставляет преподаватель в процессе защиты решения кейс-задания. Оценка выставляется по 100-балльной шкале. Оценка менее 67 баллов соответствует оценке «не зачтено».

ИТОГОВАЯ АТТЕСТАЦИЯ

Порядок проведения итоговой аттестации.

Итоговое тестирование – экзамен, включающий в себя тест и решение практико-ориентированной задачи (практического кейса)

ТЕСТ

Компьютерное тестирование включает в себя 60 тестовых вопросов, на ответы дается 90 мин, количество попыток – 3.

Примеры вопросов теста для итоговой аттестации:

1. Организационная структура, которая предполагает двойное подчинение, например, начальнику функциональному отделу и менеджеру проекта, называется

- Процессная
- Распределенная
- Проектная
- Функциональная

2. Требование “Пользовательский GUI должен предоставлять возможность языковой локализации: выбор языка (русский/английский) для надписей на элементах” – это ...

- Функциональное требование (functional requirement)
- Бизнес-требование (business requirement)
- Требование стейкхолдера (stakeholder requirement)
- Нефункциональное требование (non-functional requirement)

3. Аналог BPMN-диаграммы в

UML – это ...

- Диаграмма классов (Class diagram)
- Диаграмма компонентов (Component diagram)
- Диаграмма деятельности (activity diagram)
- Диаграмма состояний (State diagram)

4. Диаграмма Исикавы (рыбья кость) нужна, чтобы ...

▪ определить потенциальные источники проблемы и оценить степень их влияния на результат

- сформировать полный комплект документации СМК
- показать причинно-следственную связь процессов с результатом
- повысить уровень управляемости бизнес-процессов

5. Технологический процесс (workflow) отображает последовательность операций: □ рассматриваемых, как работы некоторой организационной единицы, одного или нескольких механизмов

- некоторой системы, состоящей из одного или нескольких механизмов
- рассматриваемых, как работы группы специалистов, участвующих в создании продукту процесса
- исполнителей бизнес — процесса

6. Основными требованиями к модели хозяйственной деятельности являются:

- отражение видов деятельности, допускающих стандартизацию, массовое тиражирование
- отражение наиболее существенных с точки зрения поставленной задачи видов деятельности
- максимально подробное отражение всех видов деятельности
- отражение видов деятельности, приносящих максимальный доход
- отражение видов деятельности, отвлекающих наибольший объем ресурсов

7. План можно назвать стратегией, если:

- это детальный всесторонний комплексный план, предназначенный для того, чтобы обеспечить осуществление миссии организации и достижение ее целей
- он зафиксирован в документе, сформированном при создании компании, и необходим для разъяснения ее целей представителям внешнего окружения
- это комплексный детальный план мероприятий на несколько лет вперед
- он предназначен для того, чтобы обеспечить осуществление миссии организации и достижение ее целей

8. Одним из основных элементов ABC управляемого временем является:

- номенклатура потребляемых продуктами и клиентами ресурсов
- стоимость потребляемых продуктами и клиентами ресурсов
- интенсивность потребления ресурсов продуктами и клиентами
- продолжительность потребления ресурсов продуктами и клиентами

9. Для убывающего КРІ связанного с некоторой целью следует указать:

- среднее ожидаемое за период значение
- минимально допустимое значение
- максимально допустимое значение

10. С увеличением числа компонент и связей в системе:

- можно поддерживать неизменной сложность описания системы
- описание системы постоянно усложняется
- постепенно можно использовать упрощающие предположения
- приводит к случайным колебаниям сложности описания

Порядок проведения: тестирование проводится в режиме онлайн, 20 тестовых вопросов по каждой дисциплине, 90 мин, количество попыток – 2.

Критерии оценивания теста:

Набранное количество баллов (% от максимального значения)	Оценка (баллов)
<50%	1
50% ≤, <65%	2
65% ≤, 75% <	3
75% ≤, 85% <	4
85% ≤,	5

ПРАКТИКО-ОРИЕНТИРОВАННЫЕ ЗАДАЧИ

Практико-ориентированные задачи (кейсы) для решения в рамках итоговой аттестации были предоставлены:

1. Международным евразийским индустриальным фондом
2. Инжиниринговым центром "Гипертех"
3. Центром развития базисных инноваций и прототипирования (ООО «ЦРБИП»)
4. ООО «Актив Бизнес консалтинг»
5. ООО «Вымпелком» (Билайн)
6. Общество с ограниченной ответственностью «Цифровая Формула Доверия» (ООО «Цифровая Формула Доверия»)

Кейс 1 «Оценки комплектности и качества неструктурированного (в т.ч. визуального) содержимого документов, подаваемых в составе заявки на меры государственной поддержки»

Описание задачи: по дата-сету, содержащему документы заявок на получение грантов в формате pdf, создать модель для первичной оценки содержания каждого документа заявителя и предоставления рекомендации по результату анализа:

- Представлен ли полный комплект документов;
- Соответствует ли документ по содержанию тому документу, который требовалось подать;
- Корректность оформления документов и полнота документов (наличие печатей, отметок или подтверждения получения из ФНС, ФСС, Росстат);
- Содержит ли документ необходимые реквизиты, с указанием, где каждый из них в документе находится (с визуальным указанием).

Кейс 2 «Прогнозирование числа пострадавших в ДТП на выбранном участке дорог на основе исторических данных»

Описание задачи:

Дан дата-сет, содержащий граф дорог Москвы, количество пострадавших в ДТП и информацию о перекрытиях и ремонтах дорог.

Необходимо построить модель, прогнозирующую число пострадавших в ДТП на выбранном участке дорог на основе исторических данных, а также определить:

- для заданного участка дорог спрогнозировать число пострадавших;
- выявить 10 наиболее опасных мест;
- показать, какие факторы (признаки) оказали наибольшее влияние;
- дать рекомендации, которые помогут снизить аварийность на установленном участке.

Кейс 3 «Прогнозирование нагрузки на энергосеть»

Описание: В настоящее время стремительно развиваются системы управления энергопотреблением (EMS System), позволяющие снизить плату за электрическую энергию за счет оптимального управления системой электроснабжения здания. В офисном здании установлена система накопления электрической энергии, для формирования оптимальных управляющих воздействий на которую требуется выполнение качественного прогнозирования электрической нагрузки. Это позволяет задействовать накопители электрической энергии в нужные периоды времени для снижения потребляемой из сети электрической энергии в часы пиковой нагрузки (при использовании двухставочного тарифа) или в пиковой зоне суток (при использовании дифференцированного тарифа на электрическую энергию). Более высокое качество прогноза потребления электрической энергии позволяет более эффективным образом задействовать накопитель

электрической энергии, а, следовательно, получать большую выгоду от его использования.

Необходимо подготовить модель машинного обучения, обеспечивающую высокую точность прогнозирования нагрузки офисного здания в условиях пандемии COVID-19.

Проблема: в связи с пандемией COVID-19, переход ряда сотрудников на удаленную работу приобрел стохастический характер, что требует совершенствования используемых подходов к прогнозированию потребления электрической энергии. В связи с этим необходимо учесть дополнительные параметров для прогнозирования.

Дата-сет включает данные потребления электроэнергии building_data.xlsx (с датчиков ICP DAS), данные с приборов учета building_data_meters.xlsx и метеоданные.

Кейс 4. «Работа для людей с ОВЗ»

Описание задачи: Фонд ДаМир помощи людям с ОВЗ получил список вакансий с сайта РАБОТА.ru. Необходимо провести и составить список таблиц по Московской области с возможностью выборки самых вакансий для разных категорий инвалидов по нозологии (не зрячие, слабо видящие, глухие, глухо – немые).

Задача: Ссылка на скачивание исходных ресурсов:
<https://disk.yandex.ru/d/iCRE00xUd4bTDA>

Существует таблица всех вакансий в Москве и Московской области. Необходимо продумать и сделать удобную структуру разбиения таблицы, так как отсутствуют технические мощности. Учитывать, что таблица пополняется каждый день и информация может быть разной. Список примерных задач, которые необходимо сделать:

1. Создать таблицу людей по нозологиям (категории). Необходимо парсить данные в режиме реального времени актуальный состав из ГИС «Реестр инвалидов».

2. Разбить витрину на города Московской области. Из витрины необходимо убрать ненужные столбцы (на свое усмотрение, при сдаче проекта будет дискуссия, нужно будет обосновывать). Добавить столбцы с нозологией. Убрать вакансии, не связанные с инвалидами определенных нозологий.

3. Сделать витрину и построить дашборды, которые будут показывать, сколько вакансий появляется каждый день и у каких компаний, отфильтровать от большего к меньшему.

Результат: Должна получиться таблица вакансий, в которую льются данные с ГИС, витрина с модифицированными вакансиями, несколько дашбордов.

Кейс 5 «Прогнозирование ухода клиента из банка»

Описание: из банка каждый месяц стали уходить клиенты. Маркетологи банка посчитали, что сохранять текущих клиентов дешевле, чем привлекать новых. Необходимо спрогнозировать, уйдёт ли клиент из банка в ближайшее время или нет.

Дата-сет включает исторические данные о поведении клиентов и расторжении договоров с банком.

Кейс 6 «Распознавание типов почв»

Описание: Компания занимается прогнозированием урожая сельскохозяйственных культур на основании спутниковых снимков. Проблема – облачность на снимках, которая не дает получить корректные значения вегетационных индексов. Дата-сет (case_clouds.xlsx) включает данные со значениями каналов и индексов со спутника для конкретных полей за определенный период. Необходимо подготовить модель машинного обучения, которая позволит отфильтровать снимки с облаками.

Кейс 7 «Прогнозирование количества углерода в почве»

Описание: Компания занимается прогнозированием количества углерода в почве. Необходимо построить модель, которая должна предсказывать тип почвы. Такая модель должна оптимизировать сельскохозяйственные процессы, чтобы фермер мог выбрать поле определенного типа почвы с наибольшим количеством углерода.

Дата-сеты (points_data.csv и satellites_data.csv) включают данные спутниковой съемки – значения нескольких спектральных индексов.

Кейс 8 «Фильмы к 9 мая»

Описание: выбрать наиболее интересные для зрителя фильмы о войне – чтобы добавить их в трансляцию на 9 мая.

Задача: Ссылка на скачивание исходных ресурсов: <https://disk.yandex.ru/d/3534562wuwkFi5TQ>

Существует таблица с отзывами на фильмы о войне (подборка 50 фильмов с Кинопоиска).

1. Собрать витрины, исходя из исходной таблицы, с названиями фильмов. Оставить только нужные столбцы (рейтинг, название отзыва, дату отзыва, ревью).

2. В витринах оставить только уникальные, положительные отзывы и отзывы, которые содержат в себе более 40 слов. После этого явно идентифицировать каждый отзыв.

3. Составить график, по которому можно ответить на вопрос: «Какие фильмы, по мнению людей, были самые успешные?»

Результат: должно получиться несколько витрин и один график (запрос для построения графика сохраняем).

Кейс 9 «Локомотивное депо»

Описание: Компания занимается обслуживанием локомотивов. Необходимо построить модель, которая может предсказывать значение толщины гребня колесных пар для планирования заходов локомотивов с критичными параметрами в депо и оптимизации ремонтных работ.

Дата-сет (measurements.csv) включает данные с параметрами колесных пар за определенный период.

Кейс 10 «Предсказание стоимости автомобиля»

Описание: Компании, в которой можно приобрести автомобили с пробегом, нужен сервис быстрой оценки стоимости автомобиля. Необходимо построить модель для определения стоимости автомобиля, используя данные дата-сета (autos.csv) с такими историческими данными, как технические характеристики, комплектации и цены автомобилей.

Кейс 11. «Межгосударственный сервис»

Описание: необходимо создать систему госуслуг среди дружественных стран (Белоруссия – Россия – Армения) как торговый маркетплейс для чего необходимо подготовить данные для анализа наиболее популярной продукции. Задача: Ссылка на скачивание исходных ресурсов: https://disk.yandex.ru/d/swuEI344UbS2--_ER. Берем таблицу с описанием товаров, товарами, отзывами и выполняем следующие задачи:

1. Составить витрины с разделением товаров по категориям (ввести категории).

2. В каждой витрине убрать те отзывы, которые написаны, но весь отзыв составляет менее 6 слов.

3. Для каждой витрины составить столбец и зарандомить цену для категории. Добавить столбец с датой, сгенерировать дату написания отзыва.

4. Составить график с топами товаров.

Результат: должно получиться несколько витрин и один график с лучшими товарами по мнению пользователей.

Кейс 12. «Анализ конкурентов»

Описание: необходимо спарсить отзывы о продукции конкурентов, чтобы узнать ее минусы и плюсы, выбор конкурентов и продукции.– за вами.

Задача: Необходимо написать скрипт, который будет парсить информацию с сайта (выбор сайта остается на ваш выбор) и загружать ее в любую удобную таблицу. Минимальный размер файла – 100 Мб. Минимальное количество столбцов – 20. Форматы допустимых данных – числа, текст, bool. При проверке будет проверяться количество строк, дубли информации (не должно быть файла, где одна половина идентична другой).

Результат: должен получиться файл CSV/Parquet и таблица в вашей БД.

Кейс 13. «Жизнь студента в общетии»

Описание: Необходимо выявить студентов университета, которые нуждаются в улучшении условий проживания в общежитии.

Задача: Ссылка на скачивание исходных ресурсов:
<https://disk.yandex.ru/e/dskdfC5353FVw>

В таблице содержится список комнат общежитий Университета, количество строк 734. Необходимо сделать следующие задачи:

1. Разбить витрину на номера общежитий. Убрать столбец с описаниями общежития, добавив колонку коменданта.

2. Выяснить, у каких общежитий самое большое количество нареканий на обслуживание. Сгруппировать по дате обращений, чтобы ответить на вопрос: «Какие номера чаще всего вызывают замечания?» Построить график, по которому можно понять ответ на вопрос.

Результат: должно получиться несколько витрин с различными качеством комнат и настроен мониторинг.

Кейс 14. «Аврора – российские мобильные системы»

Описание: необходимо проанализировать рынок мобильных устройств чтобы понять конкурентоспособность ОМС Автора.

Задача: Ссылка на скачивание исходных ресурсов:
<https://disk.yandex.ru/e/dskdfC5353FVw>

Существует одна большая таблица в формате CSV с отзывами. Нужно подготовить данные для аналитика с негативными, плохими, нейтральными и хорошими отзывами. Необходимо учитывать систему рейтинга.

1. Составить 4 витрины с разными отзывами.

2. К этим витринам необходимо добавить явные идентификаторы.

3. Составить список рынка мобильных устройств.

4. Так как отзыв может быть хорошим, а рейтинг ошибочным, необходимо еще раз очистить витрины – например, в витрине с хорошим рейтингом убрать те отзывы, которые содержат в себе слова «ужасно, невыносимо, некачественно, безобразно» и так далее.

Результат: Результат должен содержать в себе 4 красивые и качественные витрины, которые можно будет передать аналитику и в отдел разработки Авроры.

Кейс 15. «Импорт ОЗУ в Российскую Федерацию»

Описание: статистика импорта ОЗУ за период с 2017 по 2022 гг. Основные параметры: дата таможенных операций, код ТН ВЭД, наименование, производитель, страна происхождения, стоимость единицы, вес единицы, количество. В исходном датасете 50000 записей, представлено 22 кода ТН ВЭД, которые сгруппированы в 4 класса ОЗУ. Необходимо выделить обоснованные кластеры, на которые был спрос в течение 5 лет.

Использовать для кластеризации только 2 поля: наименование памяти и стоимость единицы. Непосредственно для кластеризации предлагается использовать несколько методов с различными гиперпараметрами. Для оценки качества кластеризации предлагается использовать метрики S_Dbw, коэффициент силуэта. Настройку гиперпараметров методов кластеризации произвести с использованием алгоритмов оптимизации гиперпараметров

(GridSearch, RandomSearch, Bayesian optimization, HyperBand, Genetic algorithm и т.п.).

В качестве результата ожидается размеченный датасет, то есть исходный датасет с добавлением поля «Метка кластера».

Кейс 16. Кейс «Детектирование наличия маски на лице»

Описание: Во время пандемии стало важно следить за тем, чтобы все надевали медицинские маски, так как от этого зависит здоровье каждого из нас. Задача состоит в обучении модели, которая умеет с высокой точностью определять, есть на лице маска или нет.

Размеченный для решения задачи дата-сет включает архивы mask.zip, содержащий 5000 квадратных кропов лиц без маски, и no_mask.zip, содержащий 4900 квадратных кропов лиц людей с маской на лице.

Кейс 17. «Видеоаналитика от Билайна»

Описание: видеонаалитику Билайна выбирают крупные компании и госструктуры. Самый популярный запрос сейчас – дефектоскопия с целью сокращения репутационных рисков за счет уменьшения бракованной продукции. Задача в MedTech & AI предсказать развитие патологии тазобедренного сустава на сервисах и дата сетах партнёрах программы.

Кейс 18. «Предложение оптимального тарифа мобильной связи»

Описание:

Известно, что многие клиенты Оператора мобильной связи используют архивные тарифы мобильной связи. Было решено построить систему для анализа поведение клиентов и рекомендации переход на новые тарифы: «Первый» или «Мой онлайн». Дата-сет, содержащий файлы users.csv, messages.csv, calls.csv и internet.csv включает данные о поведение клиентов, которые ранее перешли на эти новые тарифы. Необходимо построить модель для задачи классификации, которая выберет подходящий для клиента тариф.

Кейс 19. «Определение по фотографии возраста человека»

Описание:

Сетевой гипермаркет решил внедрить систему компьютерного зрения для обработки фотографий покупателей магазина для фотофиксации покупателей в прикассовой зоне. Собранные данные предполагается использовать для исследования совершенных покупок и предложения товаров, которые могли бы заинтересовать данную возрастную группу покупателей; контроля добросовестности кассиров при продаже покупателям алкогольной продукции. Необходимо построить модель, которая определит возраст человека по его фотографии.

Дата-сет (people_age.zip) включает фотографии людей с указанием их возраста.

Кейс 20 «Образовательный подкаст для школьников 5-6 классов»

Описание: выбрать наиболее интересные для школьников образовательные ролики о программировании – чтобы добавить их в рекомендации для школ. Задача: Ссылка на скачивание исходных ресурсов: <https://disk.yandex.ru/d/35655562wuwkFi5TQ>

Существует таблица с отзывами на подборки по программированию для школьников (подборка Яндекс класса).

1. Собрать витрины, исходя из исходной таблицы, с названиями уроков. Оставить только нужные столбцы (рейтинг, название отзыва, дату отзыва, ревью).

2. В витринах оставить только уникальные, положительные отзывы и отзывы, которые содержат в себе более 40 слов. После этого явно идентифицировать каждый отзыв.

3. Составить график, по которому можно ответить на вопрос: «Какие уроки, по мнению школьников, были самые успешные?»

Результат: должно получиться несколько витрин и один график (запрос для построения графика сохраняем).

Критерии оценивания решения кейса:

Решение кейса оценивает преподаватель, максимально возможный балл – 15.

Итоговая аттестация проводится на основании балльно-рейтинговой системы. В процессе обучения учебные действия слушателя оцениваются по шкале:

Вид учебного действия	Максимальное количество баллов
Академическая активность (посещаемость, активность на семинарах и т.д.)	15
выполнение всех практических заданий	20
текущая аттестация (по всем темам)	15
промежуточная аттестация (по всем дисциплинам)	30
Итоговая аттестация (тест + кейс)	20
ВСЕГО	100

Перевод рейтинговых баллов в оценку по традиционной четырех-балльной шкале производится в соответствии с таблицей:

Набранное количество баллов (% от максимального значения)	Оценка
<65	неудовлетворительно
65≤, 75<	удовлетворительно
75≤, 85<	хорошо
85≤,	отлично

Директор
Центра компетенций «Цифровая экономика»

**Финансового университета
при Правительстве Российской Федерации**

Сурова Н.Ю.

«21» марта 2023 г.