

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

РАССМОТРЕНО И ОДОБРЕНО

УТВЕРЖДАЮ

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

Первый проректор
С.В. Авдашкевич
28.06.2023

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина:	Б1.В.08 Анализ больших данных
Направление подготовки:	09.04.03 Прикладная информатика
Направленность (профиль):	Корпоративные информационные системы
Уровень высшего образования:	Магистратура
Форма обучения:	очная, заочная
Разработчики:	Кандидат экономических наук, доцент Щипанов Е. Ф.

Санкт-Петербург
2023

1. Цели и задачи дисциплины:

Цель освоения дисциплины:

Изучение математических методов и подходов, используемых в программных системах обработки и анализа больших данных для разработки и управления инфраструктурой с учетом технико-коммерческого предложения.

Задачи дисциплины:

Изучить основные и терминологию в области технологий обработки и анализа больших данных, понятие модели параллельной обработки данных MapReduce и ее практическую реализацию в Hadoop, основные API и типовые примеры программирования в Hadoop, стек Hadoop-технологий для распределенного хранения и обработки данных;

Развить профессиональные навыки студентов за счет получения практического опыта работы с IT решениями в части обработки и анализа больших данных.

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

Планируемые результаты освоения ОП ВО (код и содержание компетенций)	Планируемые результаты обучения по ОП ВО (индикаторы достижения компетенций)	Примечание
ПК-2 Разработка и управление инфраструктурой с учетом технико-коммерческого предложения	ПК-2.1 Знать: Компетенции и технологические возможности организации-поставщика; Возможности систем поддержки разработки и сопровождения требований; Процессы разработки и сопровождения требований.	06.017 Профессиональный стандарт «Руководитель разработки программного обеспечения»
	ПК-2.2 Уметь: Проводить презентации; Продавать идеи, услуги и решения; Управлять проектами.	
	ПК-2.3 Владеть: Проведение интервью с потенциальными клиентами; Определение потребностей и интересов потенциальных клиентов; Разработка черновых концепций системы по запросам потенциальных клиентов; Проведение экономических расчетов окупаемости предложенного варианта черновой концепции; Проведение презентации и защиты технико-коммерческого предложения; Организация выявления потребностей аналитиков и заинтересованных лиц в отношении информационно-технической инфраструктуры поддержки процессов разработки и сопровождения требований к системам; Организация разработки концепции инфраструктуры обеспечения процесса разработки и сопровождения требований к системам; Формирование заказов на закупку, внедрение, обучение и развитие инструментов и технологий разработки требований; Контроль показателей эффективности использования инфраструктуры поддержки разработки и сопровождения требований к системам.	

Планируемые результаты обучения по ОП ВО (индикаторы достижения компетенций)	Планируемые результаты обучения по дисциплине
ПК-2.1. Знать: Компетенции и технологические возможности организации-поставщика; Возможности систем поддержки разработки и сопровождения требований; Процессы разработки и сопровождения требований.	Знает компетенции и технологические возможности организации-поставщика; возможности систем поддержки разработки и сопровождения требований; процессы разработки и сопровождения требований с целью создания ПО для анализа больших данных.

Планируемые результаты обучения по ОП ВО (индикаторы достижения компетенций)	Планируемые результаты обучения по дисциплине
ПК-2.2. Уметь: Проводить презентации; Продавать идеи, услуги и решения; Управлять проектами.	Умеет проводить презентации, продавать идеи, услуги и решения; управлять проектами с целью создания ПО для анализа больших данных.
ПК-2.3. Владеть: Проведение интервью с потенциальными клиентами; Определение потребностей и интересов потенциальных клиентов; Разработка черновых концепций системы по запросам потенциальных клиентов; Проведение экономических расчетов окупаемости предложенного варианта черновой концепции; Проведение презентации и защиты технико-коммерческого предложения; Организация выявления потребностей аналитиков и заинтересованных лиц в отношении информационно-технической инфраструктуры поддержки процессов разработки и сопровождения требований к системам; Организация разработки концепции инфраструктуры обеспечения процесса разработки и сопровождения требований к системам; Формирование заказов на закупку, внедрение, обучение и развитие инструментов и технологий разработки требований; Контроль показателей эффективности использования инфраструктуры поддержки разработки и сопровождения требований к системам.	Владеет навыками разработки черновых концепций системы по запросам потенциальных клиентов; организации разработки концепции инфраструктуры обеспечения процесса разработки и сопровождения требований к системам для анализа больших данных.

3. Содержание, объем дисциплины и формы проведения занятий

№ п/п	Наименование темы дисциплины	Компетенции	Оценочные средства текущего контроля		
			ЗНАТЬ	УМЕТЬ	ВЛАДЕТЬ
			ПК-2.1	ПК-2.2	ПК-2.3
1	Технологии хранения и обработки больших данных	ПК-2	Тестирование №1 (10)	Круглый стол, дискуссия, полемика, дебаты/Эссе №1 (20)	Задания творческого уровня №1 (20)
2	Программирование обработки и загрузки больших данных в SAS	ПК-2	Тестирование №1 (10)	Круглый стол, дискуссия, полемика, дебаты/Эссе №1 (20)	Деловая и (или) ролевая игра/Кейс-задача №1 (20)
3	Аналитика в больших данных	ПК-2	Тестирование №2 (10)	Круглый стол, дискуссия, полемика, дебаты/Эссе №1 (20)	Деловая и (или) ролевая игра/Кейс-задача №1 (20)
4	Аналитическая обработка структурированных больших данных	ПК-2	Тестирование №2 (10)	Круглый стол, дискуссия, полемика, дебаты/Эссе №1 (20)	Деловая и (или) ролевая игра/Кейс-задача №2 (20)
Количество баллов (100 баллов):			100		

Содержание учебного материала, лабораторные работы и практические занятия, курсовая работа
<p>Тема 1: Технологии хранения и обработки больших данных Введение в большие данные. Практические занятия/самостоятельная работа: Модель вычислений MapReduce. Реализации алгоритмов на MapReduce. Основы Hive. Лабораторная работа: -</p>
<p>Тема 2: Программирование обработки и загрузки больших данных в SAS Основы языка SAS Base. Макропроцессор SAS. SAS SQL.</p>

Содержание учебного материала, лабораторные работы и практические занятия, курсовая работа
<p>Практические занятия/самостоятельная работа: Обзор SAS Data Loader for Hadoop. Взаимодействие SAS и Hadoop. Основы программирования на SAS Data Step 2</p> <p>Лабораторная работа: -</p> <p>Тема 3: Аналитика в больших данных Обзор задач и методов машинного обучения и интеллектуального анализа больших данных.</p> <p>Практические занятия/самостоятельная работа: Интерактивное исследование данных с помощью SAS VA/VS. Построение моделей с помощью IMSTAT в SAS LASR сервере. Интеллектуальный анализ данных с SAS Enterprise Miner. Операционализация моделей.</p> <p>Лабораторная работа: -</p> <p>Тема 4: Аналитическая обработка сложно-структурированных больших данных Обработка текстов в SAS Text Miner.</p> <p>Практические занятия/самостоятельная работа: Анализ временных рядов в SAS Forecast Studio. Анализ взаимосвязей с помощью SAS SNA. Решение оптимизационных задач в SAS/OR</p> <p>Лабораторная работа: -</p> <p>Курсовая работа: не предусмотрено учебным планом</p>

Очная форма обучения

Вид учебной работы	Всего часов	Семестр 3
Аудиторные занятия (АЗ):	36	36
Лекционные занятия (Лек)	18	18
Лабораторные занятия (Лаб)	0	0
Практические занятия (Пр)	18	18
Самостоятельная работа студента (СР)	65	65
Курсовая работа	0	0
Другие виды самостоятельной работы*	65	65
Контроль самостоятельной работы (КСР)	7	7
Контактная работа (КоР)	43	43
Форма промежуточной аттестации	0	Зачет
Подготовка к экзамену и сдача экзамена (СР, КоР)	0	0
Общая трудоемкость дисциплины, часы/ЗЕТ	108/3	108/3

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

№	Наименование темы дисциплины	Семестр Курс	Количество учебных часов				Практическая подготовка
			В том числе по видам аудиторных занятий			СР	
			Лек	Пр	Лаб		
1	Технологии хранения и обработки больших данных	3	6	6	0	17	6
2	Программирование обработки и загрузки больших данных в SAS	3	4	4	0	16	4
3	Аналитика в больших данных	3	4	4	0	16	4
4	Аналитическая обработка сложно-структурированных больших данных	3	4	4	0	16	4
Итого:			18	18	0	65	18

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

Заочная форма обучения

Вид учебной работы	Всего часов	Семестр 5
Аудиторные занятия (АЗ):	18	18
Лекционные занятия (Лек)	8	8
Лабораторные занятия (Лаб)	0	0
Практические занятия (Пр)	10	10
Самостоятельная работа студента (СР)	82	82

09.04.03 Прикладная информатика, направленность (профиль) "Корпоративные информационные системы"
 Рабочая программа дисциплины
 Дисциплина: Б1.В.08 Анализ больших данных
 Форма обучения: очная, заочная
 Разработана для приема 2020/2021, 2021/2022, 2022/2023 учебного года
 Обновлено на 2023/2024 учебный год

Вид учебной работы	Всего часов	Семестр 5
Курсовая работа	0	0
Другие виды самостоятельной работы*	82	82
Контроль самостоятельной работы (КСР)	4	4
Контактная работа (КоР)	22	22
Форма промежуточной аттестации	0	Зачет
Подготовка к экзамену/зачету и сдача экзамена/зачета (СР, КоР)	4	4
Общая трудоемкость дисциплины, часы/ЗЕТ	108/3	108/3

* Подготовка к аудиторным занятиям

№	Наименование темы дисциплины	Семестр/ Курс	Количество учебных часов				Практическая подготовка
			В том числе по видам аудиторных занятий			СР	
			Лек	Пр	Лаб		
1	Технологии хранения и обработки больших данных	5	2	4	0	21	6
2	Программирование обработки и загрузки больших данных в SAS	5	2	2	0	21	4
3	Аналитика в больших данных	5	2	2	0	20	4
4	Аналитическая обработка сложно-структурированных больших данных	5	2	2	0	20	4
Итого:			8	10	0	82	18

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

4. Способ реализации дисциплины

Без использования онлайн-курса.

5. Учебно-методическое обеспечение дисциплины:

Основная литература:

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

2. АНАЛИЗ ДАННЫХ. Учебник для вузов / Под ред. Мхитаряна В.С. - Национальный исследовательский университет «Высшая школа экономики» (г. Москва), 2022 г. - 490 с. - ISBN 978-5-534-00616-2 – Режим доступа: <https://urait.ru/book/analiz-dannyh-489100>

3. ПОСТРЕЛЯЦИОННЫЕ ХРАНИЛИЩА ДАННЫХ. Учебное пособие для вузов / Парфенов Ю. П. ; под науч. ред. Папуловской Н.В. - Уральский федеральный университет имени первого Президента России Б.Н. Ельцина (г. Екатеринбург), 2022 г. - 121 с. - ISBN 978-5-534-09837-2 – Режим доступа: <https://urait.ru/book/postrelyacionnye-hranilischa-dannyh-492609>

Дополнительная литература:

1. УПРАВЛЕНИЕ ДАННЫМИ 3-е изд., пер. и доп. Учебное пособие для вузов / Толстобров А. П. - Воронежский государственный университет (г. Воронеж), 2022 г. - 272 с. - ISBN 978-5-534-14162-7 – Режим доступа: <https://urait.ru/book/upravlenie-dannymi-496748>

2. СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ, ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА И СЛУЧАЙНЫЕ ПРОЦЕССЫ 2-е изд., испр. и доп. Учебное пособие для вузов / Берикашвили В. Ш., Оськин С. П., 2022 г. - 164 с. - ISBN 978-5-534-09216-5 – Режим доступа: <https://urait.ru/book/statisticheskaya-obrabotka-dannyh-planirovanie-eksperimenta-i-sluchaynye-processy-493106>

3. ТЕОРИЯ АЛГОРИТМОВ. ВВЕДЕНИЕ В СЛОЖНОСТЬ ВЫЧИСЛЕНИЙ 2-е изд., испр.

и доп. Учебное пособие для вузов / Крупский В. Н. - Московский государственный университет имени М.В. Ломоносова (г. Москва), 2022 г. - 117 с. - ISBN 978-5-534-04817-9 – Режим доступа: <https://urait.ru/book/teoriya-algoritmov-vvedenie-v-slozhnost-vychisleniy-492937>

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

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

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

1. ibooks.ru : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <https://ibooks.ru>. - Текст: электронный
2. Электронно-библиотечная система СПБУТУиЭ : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <http://libume.ru>. - Текст: электронный
3. Юрайт : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <https://urait.ru>. - Текст: электронный
4. [eLibrary.ru](http://elibrary.ru) : научная электронная библиотека [Электронный ресурс] : профессиональная база данных. - Режим доступа: <http://elibrary.ru>. - Текст: электронный
5. Архив научных журналов НЭИКОН [Электронный ресурс] : профессиональная база данных. - Режим доступа: agch.neicon.ru. - Текст: электронный
6. КиберЛенинка : научная электронная библиотека [Электронный ресурс] : информационная справочная система. - Режим доступа: <http://cyberleninka.ru>. - Текст: электронный
7. Лань : электронно-библиотечная система [Электронный ресурс] : профессиональная база данных. - Режим доступа: <https://e.lanbook.com>. - Текст: электронный
8. [it-world.ru](https://www.it-world.ru) [Электронный ресурс] : информационная справочная система . - Режим доступа: <https://www.it-world.ru>. - Текст: электронный
9. Виртуальный компьютерный музей [Электронный ресурс] : информационная справочная система . - Режим доступа: <https://www.computer-museum.ru>. - Текст: электронный
10. Компьютерра : информационная справочная система . - Режим доступа: <https://www.computerra.ru/>. - Текст: электронный
11. Connect: IT-технологии : информационная справочная система. - Режим доступа: <https://www.connect-wit.ru/>. - Текст: электронный
12. Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации: профессиональная база данных. - Режим доступа: <https://digital.gov.ru>. - Текст: электронный
13. Федеральная служба по надзору в сфере связи, информационных технологий и массовых коммуникаций: профессиональная база данных . - Режим доступа: <https://rkn.gov.ru>. - Текст: электронный
14. [Math-Net.Ru](https://www.mathnet.ru/): профессиональная база данных . - Режим доступа: <https://www.mathnet.ru/>. - Текст: электронный

8. Материально-техническое обеспечение дисциплины

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

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

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

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

9. Оценочные материалы по дисциплине

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

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

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

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

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

Шкала оценивания учебных достижений по дисциплине, завершающейся зачетом без оценки

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

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

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

**9.1. Типовые контрольные задания для текущего контроля
Тестирование №1**

1. Аналитик это ...

- а) специалист в области анализа и моделирование
- б) 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 характеристикам 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 характеристикам 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 характеристикам 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 характеристикам 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 анализа.

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 анализа.

11 Регрессивные модели 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.

12. Виды лингвистической неопределенности:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами;
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
- в) случайность (или наличие во внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов

(многозначность, размытость, неясность, нечеткость)

г) неоднозначность смысла фраз (Синтаксическая и семантическая).

13. Модели исключений описывают ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

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

14 Итоговые модели обнаружат ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

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

15 Модели ассоциации проявляют ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

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

Тестирование №2

1. Виды физической неопределенности данных:

а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие во внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью)

б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);

в) случайность (или наличие во внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);

г) неоднозначность смысла фраз (Синтаксическая и семантическая).

2. Очистка данных — ...

а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т. д.

б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач

в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитического задачи

г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

3. Обогащение — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

4. Консолидация — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

5. Транзакция — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных
- б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов
- в) высокоуровневые средства отражения информационной модели и описания структуры данных
- г) это установление зависимости дискретной выходной переменной от входных переменных

6. Метаданные — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных
- б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов
- в) высокоуровневые средства отражения информационной модели и описания структуры данных
- г) это установление зависимости дискретной выходной переменной от входных переменных

7. Классификация — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано

с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных

8. Регрессия — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

9. Кластеризация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных.

10. Ассоциация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

11. Машинное обучение — ...

а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.

г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

12. Аналитическая платформа — ...

а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему

правильный выходной результат.

г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

13. Обучающая выборка — ...

а) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

б) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему

правильный выходной результат.

г) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

14. Ошибка обучения — ...

а) это ошибка, допущенная моделью на учебной множества.

б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества

в) имена, типы, метки и назначения полей исходной выборки данных

г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

15. Ошибка обобщения — ...

а) это ошибка, допущенная моделью на учебной множества.

б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества

в) имена, типы, метки и назначения полей исходной выборки данных

г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

Круглый стол, дискуссия, полемика, дебаты/Эссе №1

1. Анализ временных рядов в SAS Forecast Studio.
2. Анализ взаимосвязей с помощью SAS SNA.
3. Обзор SAS Data Loader for Hadoop.
4. Взаимодействие SAS и Hadoop.
5. Интерактивное исследование данных с помощью SAS VA/VS.
6. Интеллектуальный анализ данных с SAS Enterprise Miner.

Задание творческого уровня №1

Дан текстовый корпус документов, лежащих в указанной директории. Создайте в SAS Text Miner проект, который: выберет файлы с расширением pdf; осуществит парсинг набора с определением частей речи и сохранением в признаковом пространстве только существительных и глаголов; осуществит фильтрацию документов и признаков с использованием заданной схемой определения весов лексем (например, на основе tf-idf); выделит заданное количество ключевых тематик по методу SVD. В ответе укажите топ 5 ключевых слов во второй выявленной тематике. Какой документ имеет наибольший вес в это тематике?

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

Дан набор данных заданной структуры и программа SAS Data step, производящая определенную обработку и вычисления с использованием данного набора. Перепишите эту

программу на SAS DS2 с использованием параллельных нитей и созданием пользовательского пакета, чтобы результат обработки сохранился тем же, но код мог выполняться в параллельной

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

Дан набор заданной структуры, постройте модель прогнозирования от-клика с использованием процедуры `impstat` с алгоритмом `random forest` с заданным числом деревьев. Примените полученную модель к тестовому набору данных той же структуры, визуализируйте полученный график Lift. Постройте на том же наборе модель с использованием высокопроизводительной версии метода GLM. Примените к тестовому набору. Сравните результаты GLM и Random Forest по AUC.

9.2. Примерный перечень тем курсовой работы

Не предусмотрено учебным планом

9.3. Типовые контрольные задания для промежуточной аттестации: зачет

Примерный перечень теоретических вопросов к зачету

1. Большие данные, основные понятия и определения
2. Основные проблемы больших данных и пути их решения
3. Технология больших данных
4. Роль больших данных в технике
5. Роль больших данных в экономике
6. Роль больших данных в производстве
7. Примеры и проблемы использования больших данных
8. Техники больших данных. Общие понятия
9. Техники больших данных – консолидация данных
10. Визуализация данных
11. Кластеризация данных
12. Регрессионный анализ и большие данные
13. Техника анализа ассоциативных правил
14. Нейронные сети
15. Технологии и инструменты больших данных. Общие понятия
16. Apache Hadoop, назначение и принципы работы
17. Система потоковой обработки Storm
18. Статистическая обработка данных
19. Язык программирования R
20. Аналитика больших данных. Общие понятия
21. Жизненный цикл проекта аналитики больших данных
22. Исследование аналитики больших данных
23. Подготовка данных при аналитике
24. Планирование модели больших данных
25. Построение модели больших данных
26. Использование модели больших данных

Примерный перечень практических заданий к зачету

Решить задачу: имеется csv-лог рекламной системы вида:

1111,RU,Moscow,2,4,0.3

2222,RU,Voronezh,2,3,0.2

13413,UA,Kiev,4,11,0.7

Необходимо рассчитать среднюю стоимость показа рекламы по городам России с использованием MapReduce.

09.04.03 Прикладная информатика, направленность (профиль) "Корпоративные информационные системы"

Рабочая программа дисциплины

Дисциплина: Б1.В.08 Анализ больших данных

Форма обучения: очная, заочная

Разработана для приема 2020/2021, 2021/2022, 2022/2023 учебного года

Обновлена на 2023/2024 учебный год
