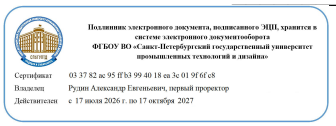
Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Санкт-Петербургский государственный университет промышленных технологий и дизайна»
(СПбГУПТД)

УТВЕРЖДАЮ
Первый проректор, проректор
по УР
_____А.Е. Рудин



Рабочая программа дисциплины

Б1.В.14 Основы анализа данных и Data Mining

Учебный план: 2026-2027 09.03.01 ВШПМ Разр IT-сист и мультим прил ОО №1-1-55.plx

Кафедра: 21 Информационных и управляющих систем

Направление подготовки: (специальность) 09.03.01 Информатика и вычислительная техника

Профиль подготовки: (специализация) Разработка IT-систем и мультимедийных приложений

Уровень образования: бакалавриат

Форма обучения: очная

План учебного процесса

Семестр (курс для ЗАО)		Контактная работа обучающихся		Сам. работа	Контроль, час.	Трудоёмкость, ЗЕТ	Форма промежуточной аттестации
		Лекции	Практ. занятия				
7	УП	32	32	51	29	4	Экзамен, Курсовая работа
	РПД	32	32	51	29	4	
Итого	УП	32	32	51	29	4	
	РПД	32	32	51	29	4	

Санкт-Петербург
2026

Рабочая программа дисциплины составлена в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 09.03.01 Информатика и вычислительная техника, утверждённым приказом Минобрнауки России от 19.09.2017 г. № 929

Составитель (и):

кандидат технических наук, Доцент	_____	Белая Т.И.
От кафедры составителя: Заведующий кафедрой информационных и управляющих систем	_____	Дроздова Елена Николаевна
От выпускающей кафедры: Заведующий кафедрой	_____	Дроздова Елена Николаевна

Методический отдел:

_____

1 ВВЕДЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

1.1 Цель дисциплины: Сформировать у студентов компетенции в области применения технологий обработки данных (в том числе big data) и машинного обучения к решению прикладных задач, современных проблем прикладной математики и информатики, проблем обработки и анализа информации.

1.2 Задачи дисциплины:

− формирование у студентов представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining).

− изучение основных подходов и алгоритмов решения задач анализа данных и особенностей их применения к решению реальных задач.

− получение студентами навыка по выявлению, формализации и успешному решению практических задач анализа данных, возникающие в процессе их профессиональной деятельности.

− получение практического навыка в работе с существующими программными пакетами по анализу данных.

1.3 Требования к предварительной подготовке обучающегося:

Предварительная подготовка предполагает создание основы для формирования компетенций, указанных в п. 2, при изучении дисциплин:

Программирование

Философия

Основы системного анализа

Алгоритмы и структуры данных

Правоведение

Теория информации

Вычислительная математика

Языки и методы программирования

Методы программирования

Системы искусственного интеллекта

Имитационное моделирование

2 КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-1: Способен осуществлять проектирование и дизайн информационных систем
Знать: теоретические и методологические основы интеллектуального анализа; методики предобработки первичных данных; особенности методов интеллектуального анализа данных; особенности методов визуального интеллектуального анализа
Уметь: выбирать средства анализа, наиболее эффективные для конкретных данных; применять методы первичной обработки данных; правильно понимать и интерпретировать полученные результаты исследования
Владеть: навыками сбора первичной организации и хранения данных для конкретного исследования; навыками самостоятельного проведения исследований

3 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Наименование и содержание разделов, тем и учебных занятий	Семестр (курс для ЗАО)	Контактная работа		СР (часы)	Форма текущего контроля
		Лек. (часы)	Пр. (часы)		
Раздел 1. Технологии анализа данных	7				О
Тема 1. Введение. Понятие анализа данных. Задачи систем поддержки принятия решений. OLTP и OLAP-системы. Принципы построения информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Правила Кодда. Размерностные модели. MOLAP, ROLAP, HOLAP-системы. Витрины данных. ETL (ExtractingTransformatingandLoading) – средство извлечения, обработки и загрузки данных. Добыча данных. Добыча данных в управлении качеством. DataMining. Стандарты DataMining. Стандарт CWM, CRISP, PMML. Жизненный цикл процесса анализа данных. Классификация методов DataMining. Модели DataMining. Понятие данные и знания. Процесс обнаружения знаний. Классификация задач DataMining. Методы анализа данных. Разведочный анализ данных. Очистка и фильтрация данных. Статистические диаграммы. «Ящичные» диаграммы. Диаграммы «ствол-листья». Задачи классификации и регрессии. Использование статистических пакетов для интеллектуального анализа данных. Понятие бизнес-аналитики. Средства бизнес-аналитики. Средства легкой бизнес-аналитики. QlikView, QlikSence.		2		8	
Тема 2. Методология KDD. Задачи предобработки данных. Технология ETL. Просмотр данных. Очистка данных. Оценка качества данных. Заполнение пропущенных данных. Аномальные и предельные данные. Использование ящечной диаграммы. Выявление дубликатов и противоречий. Корреляционный анализ. Использование факторного анализа при предобработке данных. Трансформация данных. Квантование. Сэмплинг. Группировка данных.		4		8	
Раздел 2. Интеллектуальный анализ данных					О
Тема 3. Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Дендограммы. Метод К-средних. Профили кластеров. Взаимосвязь кластерного и регрессионного анализа. Использование пакета Deductor для решения задач кластерного анализа. Кластерный анализ в средствах интеллектулаьного анализа MicrosoftOffice. Практическое занятие 1: Подготовка данных к анализу и обработке		8	8	8	

Тема 4. Основные положения непараметрической и нечисловой статистики. Таблицы сопряженности. Таблица сопряженности 2x2. Таблицы флагов и заголовков. Непараметрические и нечисловые критерии. Канонический анализ. Корреляционная матрица. Коэффициенты канонической корреляции. Меры избыточности переменных. Задачи ассоциации. Ассоциативные правила. Поддержка и достоверность ассоциативных правил. Лифт. Алгоритмы построения ассоциативных правил. Рекомендации по генерации правил. Алгоритм apriori. Использование пакета Deductor для построения ассоциативных правил. Практическое занятие 2: Решение задач корреляционного анализа		8	6	10	
Раздел 3. Проверка статистических гипотез					
Тема 5. Ошибки первого и второго рода. Уровень значимости и мощность критерия. Описание гипотез и критерии их проверки. Простые и сложные гипотезы. Проверка гипотез о равенстве средних и дисперсий двух нормально распределенных генеральных совокупностей. Хи-квадрат критерий Пирсона: проверка гипотезы о соответствии наблюдаемых значений предполагаемому распределению вероятностей (дискретному или непрерывному). Проверка гипотез о вероятностной природе данных (стационарности, нормальности, независимости, однородности). Практическое занятие 3: Проверка гипотез		6	8	7	О,Пр
Тема 6. Функциональная и статистическая зависимости. Корреляционная таблица. Групповые средние. Понятие корреляционной зависимости. Эмпирическая ковариация. Выборочный коэффициент корреляции, его свойства. Основные задачи теории корреляции: определение формы и оценка тесноты связи. Виды корреляционной связи (парная и множественная, линейная и нелинейная). Линейная корреляция. Уравнения прямых регрессии для парной корреляции. Определение параметров прямых регрессии методом наименьших квадратов. Значимость коэффициентов по критерию Стьюдента. Практическое занятие 4: Проверка зависимостей		4	10	10	
Итого в семестре (на курсе для ЗАО)		32	32	51	
Консультации и промежуточная аттестация (Экзамен, Курсовая работа)		4,5		24,5	
Всего контактная работа и СР по дисциплине		68,5		75,5	

4 КУРСОВОЕ ПРОЕКТИРОВАНИЕ

4.1 Цели и задачи курсовой работы (проекта): Целями написания курсового проекта являются: закрепление и углубление знаний по анализу данных, полученных студентами в рамках изучаемой дисциплины; формирование умений применять теоретические знания при решении конкретных практических задач; приобретение и закрепление навыков самостоятельной работы.

4.2 Тематика курсовой работы (проекта): 1. Анализ продаж в розничной сети с использованием временных рядов

2. Кластеризация клиентов интернет-магазина для персонализированного маркетинга
3. Обнаружение мошенничества в банковских транзакциях с помощью машинного обучения
4. Анализ социальных медиа для предсказания трендов в общественном мнении
5. Прогнозирование цен на недвижимость с использованием регрессионных моделей
6. Анализ медицинских данных для предсказания рисков заболеваний
7. Оптимизация логистики доставки с помощью анализа больших данных
8. Изучение влияния климатических изменений на урожайность сельскохозяйственных культур

9. Анализ пользовательского поведения в мобильных приложениях

10. Прогнозирование энергопотребления в умных городах

4.3 Требования к выполнению и представлению результатов курсовой работы (проекта):

Структурными элементами КР являются:

• Титульный лист;
• Задание на КР;
• Резюме и ключевые слова;
• Содержание;
• Введение;
• Основная часть;
• Заключение;
• Список использованных источников;
• Приложения.

Титульный лист КР оформляется по образцу, предоставляемому кафедрой.

Введение должно содержать оценку современного состояния решаемой научно-технической проблемы, основание и исходные данные для разработки темы, сведения о планируемом научно-техническом уровне разработки, о патентных исследованиях и выводы из них. Во введении должны быть показаны актуальность и новизна темы, связь данной работы с другими научно-исследовательскими работами.

В основной части приводят данные, отражающие сущность, методику и основные результаты выполненной КР.

Основная часть должна содержать:

• выбор направления исследований, включающий обоснование направления исследования, методы решения задач и их сравнительную оценку, описание выбранной общей методики проведения КР;

• процесс теоретических и (или) экспериментальных исследований, включая определение характера и содержания теоретических исследований, методы исследований, методы расчета, обоснование необходимости проведения экспериментальных работ, принципы действия разработанных объектов, их характеристики;

• обобщение и оценку результатов исследований, включающих оценку полноты решения поставленной задачи и предложения по дальнейшим направлениям работ, оценку достоверности полученных результатов и технико-экономической эффективности их внедрения и их сравнение с аналогичными результатами отечественных и зарубежных работ, обоснование необходимости проведения дополнительных исследований, отрицательные результаты, приводящие к необходимости прекращения дальнейших исследований характер.

Заключение должно содержать:

• краткие выводы (в прошедшем времени, не менее 5-6 выводов) по результатам КР;
• оценку полноты решений поставленных задач;

Правила оформления КР:

Пояснительная записка должна быть напечатана компьютерным способом на одном стороне листа белой бумаги форматом 210х297 (формат А4). Опечатки, орфографические ошибки и помарки не допускаются. В пояснительной записке должны быть соблюдены размеры полей: левое не менее 30 мм, правое не менее 10 мм, верхнее и нижнее не менее 20 мм. Кегль шрифта 14 пт, интервал между строками полуторный.

Для основного текста пояснительной записки необходимо использовать выравнивание по формату, абзацный отступ не менее 1,25 см, включить запрет висячих строк. Колонцифра ставится в нижнем поле страницы и выравнивается по центру. Титульный лист входит в нумерацию, но номер страницы на нем не ставится.

Наименование структурных элементов пояснительной записки выделяют прописным шрифтом и выравнивают по центру строки. Структурные элементы пояснительной записки должны начинаться с новой страницы.

Разделы, заголовки и подзаголовки пояснительной записки выравнивают по линии абзацного отступа. Подчеркивать заголовки нельзя.

Нумерация заголовков сквозная, обозначается арабскими цифрами. Точка ставится только между номерами, между номером и текстом не ставится. Точка в конце заголовков не ставится. Слова “Глава” и “Раздел” не пишутся.

Если раздел или подраздел содержит всего один заголовок, то он не нумеруется.

Приложения нумеруются русскими или латинскими буквами. Номера страниц приложения общая с пояснительной запиской, сквозная. Каждое приложение должно начинаться с новой страницы.

Иллюстрации нумеруют арабскими цифрами сквозной нумерацией, за исключением иллюстраций приложений. Допускается нумерация в пределах раздела. Иллюстрации приложений нумеруются отдельно. Подрисуночную подпись располагают по центру строки под иллюстрацией.

Таблицы нумеруют арабскими цифрами сквозной нумерацией, за исключением иллюстраций приложений. Допускается нумерация в пределах раздела. Наименование таблиц располагают в одну строку с ее номером, над таблицей и выравнивают по левой границе таблицы без абзацного отступа.

На все иллюстрации и таблицы должны быть ссылки в тексте.

Формулы нумеруют арабскими цифрами сквозной нумерацией, за исключением формул приложений выравнивают по правой границе текста. Сверху и снизу текста должен быть интервал не менее одной строки. Ссылки на формулу помещают в круглых скобках.

При оформлении списков следует использовать тире или строчные русские буквы в алфавитном порядке, начиная с “а”. После буквы ставится скобка.

Ссылки на использованные источники указываются в квадратных скобках. Нумерация ссылок арабскими цифрами, сквозная, в порядке указания источников в списке.

Оформление списка литературы должно соответствовать ГОСТ Р 7.0.100– 2018

Студенту предоставляется слово для доклада (время доклада – 5 мин). Приветствуется научный стиль изложения, лаконизм и содержательность выводов по работе. В докладе должны быть отражены следующие основные моменты: - цель и задачи работы; - обоснование выбора языка и среды программирования; - изложение основных результатов работы; - краткие выводы по тем результатам работы, которые определяют ее практическую значимость, степень и характер новизны элементов. Доклад может сопровождаться презентацией (MS PowerPoint). После доклада студенту-автору работы задаются вопросы. Докладчику может быть задан любой вопрос по содержанию работы. Общая длительность защиты одной работы – не более 15 минут. Оценка за курсовой проект ставится с учетом: соответствия работы заданию; новизны результатов работы; практической значимости результатов работы; качества оформления; качества защиты работы студентом. Оценками курсового проекта могут быть: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

5 БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА ОЦЕНИВАНИЯ

Использование балльно-рейтинговой системы не предусмотрено.

6. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ, КРИТЕРИЕВ, СИСТЕМЫ И СРЕДСТВ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

6.1 Показатели оценивания

Код компетенции	Показатели оценивания результатов обучения
ПК-1	1. Называет основные понятия и виды интеллектуального анализа данных, формулирует основные принципы организации анализа данных, характеризует свойства исходных данных 2. Определяет вид анализа для конкретного исследования, подготавливает данные для исследования, формулирует гипотезу исследования 3. Использует специализированное программное обеспечения для решения конкретной задачи анализа данных, интерпретирует полученные результаты

6.2 Система и критерии оценивания

Шкала оценивания	Критерии оценивания сформированности компетенций	
	Устное собеседование	Письменная работа
5 (отлично)	Полный, исчерпывающий ответ, явно демонстрирующий глубокое понимание предмета и широкую эрудицию в оцениваемой области, умение использовать теоретические знания для решения практических задач.	Работа выполнена в указанные преподавателем сроки, пояснительная записка оформлена в соответствии с требованиями, недочетов в оформлении нет. Работа написана грамотным русским языком.
4 (хорошо)	Ответ полный и правильный, основанный на проработке всех обязательных источников информации. Подход к материалу ответственный, но допущены в ответах несущественные ошибки, которые устраняются только в результате собеседования Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации. Присутствуют небольшие пробелы в знаниях или несущественные ошибки.	Работа выполнена в указанные преподавателем сроки, пояснительная записка оформлена в соответствии с требованиями, содержит незначительные погрешности в оформлении Работа написана грамотным русским языком.
3 (удовлетворительно)	Ответ воспроизводит в основном только лекционные материалы, без самостоятельной работы с рекомендованной литературой. Демонстрирует понимание предмета в целом при неполных, слабо аргументированных ответах. Присутствуют неточности в ответах, пробелы в знаниях по некоторым темам, существенные ошибки, которые могут быть найдены и частично устранены в результате собеседования Ответ неполный, основанный только на	Работа выполнена позже указанных преподавателем сроков, пояснительная записка оформлена в соответствии с требованиями, содержит значительное количество небольших недочётов. Содержание пояснительной записки выполнено формально, низкий уровень уникальности. Работа написана грамотным русским языком.

	лекционных материалах. При понимании сущности предмета в целом – пробелы в знаниях сразу по нескольким темам, существенные ошибки, устранение которых в результате собеседования затруднено.	
2 (неудовлетворительно)	Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные существенные ошибки. Непонимание заданного вопроса. Неспособность сформулировать хотя бы отдельные концепции дисциплины. Попытка списывания, использования неразрешенных технических устройств или пользования подсказкой другого человека (вне зависимости от успешности такой попытки).	Работа не выполнена. Работа выполнена позже установленных сроков, оформлена с грубыми нарушениями требований, уровень уникальности предельно низкий. Наличие ошибок в тексте пояснительной записки. Работа списана.

6.3 Типовые контрольные задания и/или иные материалы, используемые для оценки результатов обучения

6.3.1 Типовые контрольные вопросы

№ п/п	Формулировки вопросов
Семестр 7	
1	Понятие Большие данные. Роль цифровой информации в 21 веке. Проблемы анализа и обработки большого объема данных.
2	Базовые принципы обработки больших данных.
3	Определение модели. Свойства модели.
4	Аналитический подход к моделированию.
5	Информационный подход к моделированию.
6	Лица, участвующие в информационном моделировании. Общая схема анализа.
7	Определение тиражирования знаний. Процесс построения модели.
8	Технологии обработки больших данных: NoSQL, MapReduce, Hadoop, R.
9	Методика извлечения знаний Knowledge Discovery in Databases (KDD). Этапы KDD.
10	Data Mining. Постановка основных задач.
11	Машинное обучение. Бизнес-решения с помощью алгоритмов Data Mining.
12	Классификация ПО в области Data Mining и KDD. Типовая схема системы на базе аналитической платформы.
13	Формальная постановка задачи кластеризации. Цели кластеризации.
14	Основные шаги алгоритма k-means. Условие остановки алгоритма k-means. Преимущества и недостатки алгоритма k-means.
15	Кластеризация с помощью самоорганизующейся карты Кохонена
16	Этапы проведения классификации. Обзор методов классификации и регрессии.
17	Задачи линейной и логистической регрессии.
18	Определение дерева решений. Структура дерева решений. Выбор атрибута разбиения в узле.
19	Алгоритм ID3. Алгоритм C4.5.

6.3.2 Типовые тестовые задания

Какая из следующих задач является примером задачи классификации в многомерном анализе данных?
Регрессия
Кластерный анализ
Ассоциативные правила
Анализ временных рядов

Что представляет собой метод главных компонент в многомерном анализе данных?
Функция распределения

Матрица ковариаций
Линейное преобразование признаков
Метод k-средних

Какие метрики могут использоваться для измерения качества кластеризации в многомерном анализе данных?
Коэффициент корреляции
Индекс силуэта
Критерий Акаике
Метод наименьших квадратов

Что такое переобучение в контексте многомерного анализа данных?
Недостаточное количество данных
Недостаточная размерность признаков
Слишком сложная модель
Слишком точное соответствие обучающим данным

Какой метод используется для заполнения пропущенных значений в многомерном анализе данных?
ANOVA
Интерполяция
Метод опорных векторов
Кросс-валидация

Что такое «бэггинг» в машинном обучении и многомерном анализе данных?
Алгоритм опорных векторов
Метод главных компонент
Ансамбль моделей, обученных на случайных подвыборках данных
Критерий информативности Джини

Какая из следующих задач является примером задачи регрессии в многомерном анализе данных?
Кластеризация текстовых документов
Определение темы новостей
Предсказание цены акций
Распознавание образов на изображениях

Что такое «однородность» в контексте кластерного анализа в многомерном анализе данных?
Количество признаков в данных
Степень корреляции между признаками
Схожесть объектов внутри кластера
Энтропия данных

Какая метрика может быть использована для оценки производительности модели классификации в многомерном анализе данных?
Коэффициент детерминации
Точность (Accuracy)
Коэффициент корреляции Пирсона
Коэффициент Жаккара

Что представляет собой метод k-средних в многомерном анализе данных?
Логистическая регрессия
Градиентный спуск
Алгоритм кластеризации, разделяющий данные на k кластеров
Метод опорных векторов

Какие из перечисленных методов являются основными при проведении многомерного анализа данных?
Метод главных компонент
Метод кратчайших квадратов
Метод случайных лесов
Метод иерархической кластеризации

Что представляют собой собственные значения в контексте многомерного анализа данных?
Числа, характеризующие количество наблюдений в выборке
Оценки значимости переменных в модели
Корни характеристического уравнения, связанного с матрицей данных
Меры сходства между наблюдениями в многомерном пространстве

Какие проблемы могут возникнуть при многомерном анализе данных из-за мультиколлинеарности?
Сложность интерпретации результатов
Переобучение модели

Неустойчивость оценок параметров
Высокая чувствительность к малым изменениям в данных

Что такое метод t-стохастического вложения (t-SNE) в контексте многомерного анализа данных?
Метод для выделения аномалий в данных
Метод для подсчета статистических t-критериев
Метод для визуализации высокоразмерных данных в двумерном пространстве
Метод для решения задачи классификации

Какие из перечисленных метрик расстояния часто используются в многомерном анализе данных?
Манхэттенское расстояние
Евклидово расстояние
Косинусное расстояние
Все вышеперечисленные метрики могут использоваться

Какие виды многомерного анализа данных используются для выделения латентных структур в данных?
Кластерный анализ
Дискриминантный анализ
Многомерный метод шкалирования
Факторный анализ

Какая задача решается с использованием метода иерархической кластеризации в многомерном анализе данных?
Понижение размерности данных
Классификация объектов
Определение важности переменных
Группировка наблюдений на основе их сходства

Что такое «скрытая переменная» в контексте факторного анализа?
Переменная, наблюдаемая в данных
Переменная, влияющая на результаты исследования
Переменная, содержащая выбросы
Латентная переменная, не наблюдаемая напрямую, но объясняющая вариабельность в данных

Что представляют собой коэффициенты факторов в факторном анализе?
Оценки значимости переменных в модели
Веса, отражающие вклад каждой переменной в формирование фактора
Коэффициенты регрессии
Множители Лагранжа

Какое из перечисленных ниже является основным этапом в процессе многомерного анализа данных?
А) Сбор данных
Б) Визуализация данных
В) Оценка статистической значимости
Г) Применение машинного обучения

Какое понятие используется для обозначения явления, при котором переменные в наборе данных взаимно связаны и могут быть представлены как матрица ковариаций?
А) Нормализация данных
Б) Декомпозиция данных
В) Многомерная корреляция
Г) Регрессионный анализ

Какие методы многомерного анализа широко используются для уменьшения размерности данных?
А) Агрегирование данных
Б) Факторный анализ
В) Дескриптивная статистика
Г) Плотностная оценка

Что представляет собой метод главных компонент в многомерном анализе данных?
А) Алгоритм кластеризации
Б) Функция плотности распределения
В) Статистическая гипотеза
Г) Линейное преобразование переменных

Какой из следующих методов подходит для анализа категориальных переменных в многомерном пространстве?
А) t-критерий Стьюдента
Б) Анализ частотных таблиц

В) Корреляционный анализ
Г) Метод наименьших квадратов

Что означает термин «собственное значение» в контексте многомерного анализа данных?
А) Число, характеризующее величину изменения при преобразовании данных
Б) Статистическая вероятность
В) Доля объясненной дисперсии
Г) Коэффициент корреляции

Какой метод используется для выявления групп схожих наблюдений в многомерном пространстве?
А) Анализ дисперсии
Б) Метод главных компонент
В) Кластерный анализ
Г) Многомерная регрессия

Какие метрики могут быть использованы для измерения расстояния между наблюдениями в многомерном пространстве?
А) Дисперсия
Б) Ковариация
В) Евклидово расстояние
Г) Косинусное сходство

Какие из перечисленных методов используются для проверки гипотезы о структуре факторов в многомерном анализе?
А) Критерий Колмогорова-Смирнова
Б) Критерий Бартлетта
В) Критерий Стьюдента
Г) Критерий Манна-Уитни

Какой из следующих методов является статистическим методом многомерного анализа данных?
Факторный анализ
Метод наименьших квадратов
Кластерный анализ
Дисперсионный анализ

Что из перечисленного представляет собой метод уменьшения размерности в многомерном пространстве?
Гипотезы о равенстве средних
Кластерный анализ
Метод главных компонент
Метод наименьших квадратов

Какой индекс можно использовать для измерения схожести между двумя наборами данных?
ANOVA
Критерий Вилкоксона
Коэффициент корреляции
Метод главных компонент

Какой метод многомерного анализа данных чаще всего используется для выделения групп схожих объектов?
ANOVA
Дисперсионный анализ
Кластерный анализ
Т-тест

Какой из перечисленных методов подходит для оценки влияния одной переменной на несколько зависимых переменных?
ANOVA
Метод главных компонент
Канонический анализ
Критерий Вилкоксона

Какой статистический метод используется для проверки гипотезы о различии средних значений в трех или более группах?
ANOVA
Метод главных компонент
Множественная регрессия
Дисперсионный анализ

Какой из методов оценивает, насколько хорошо модель соответствует данным и предсказывает новые

наблюдения?

ANOVA
Метод наименьших квадратов
Кросс-валидация
Коэффициент корреляции

Какой инструмент можно использовать для визуализации многомерных данных и выявления паттернов?

ANOVA
Метод главных компонент
Кластерный анализ
Многомерное шкалирование

Какой метод многомерного анализа данных используется для анализа связей между категориальными переменными?

ANOVA
Множественная регрессия
Анализ таблиц сопряженности
Кластерный анализ

Какая из следующих задач может быть решена с использованием метода кластерного анализа?
Прогнозирование временных рядов
Анализ связей между переменными
Группировка схожих объектов
Оценка влияния одной переменной на другую

6.3.3 Типовые практико-ориентированные задания (задачи, кейсы)

1 Найдите в сети Интернет два сайта, на которых используются системы прогнозирования.

2 Найдите в сети Интернет два сайта, на которых используются рекомендательные системы.

3 Пользуясь системой SCOPUS, проанализируйте динамику количества публикаций за пять лет по направлениям Deep Learning, Big Data, Recommender Systems, Social Network Analysis.

4 Пользуясь системой SCOPUS, найдите пять публикаций с наибольшей цитируемостью публикаций за последние десять лет по направлениям Deep Learning, Big Data, Recommender Systems, Social Network Analysis.

5 Пользуясь системами SCOPUS, Web of Science, E-library (РИНЦ), выявите нескольких ведущих ученых в сфере анализа данных.

6.4 Методические материалы, определяющие процедуры оценивания результатов обучения

6.4.1 Условия допуска обучающегося к промежуточной аттестации и порядок ликвидации академической задолженности

Проведение промежуточной аттестации регламентировано локальным нормативным актом СПбГУПТД «Положение о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся»

6.4.2 Форма проведения промежуточной аттестации по дисциплине

Устная	+	Письменная	+	Компьютерное тестирование		Иная	

6.4.3 Особенности проведения промежуточной аттестации по дисциплине

При проведении экзамена время, отводимое на подготовку к ответу, составляет не более 40 мин. Для выполнения практического задания обучающему предоставляется необходимая справочная информация.

Время, отводимое на защиту курсовой работы не более 20 минут.

Сообщение результатов обучающемуся производится непосредственно после устного ответа.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

7.1 Учебная литература

Автор	Заглавие	Издательство	Год издания	Ссылка
7.1.1 Основная учебная литература				
Нестеров, С. А.	Интеллектуальный анализ данных средствами MS SQL Server 2008	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа	2023	https://www.iprbookshop.ru/131496.html

Васильев, Е. П., Орешков, В. И.	Интеллектуальный анализ данных в технологиях принятия решений	Рязань: Рязанский государственный радиотехнический университет	2023	https://www.iprbookshop.ru/134854.html
Глушань, В. М., Норкин, О. Р., Парфенова, С. С.	Основы системного анализа. В 2 частях. Ч.2	Ростов-на-Дону, Таганрог: Издательство Южного федерального университета	2023	https://www.iprbookshop.ru/135658.html
Дятлов, А. В., Лукичев, П. Н.	Методы многомерного статистического анализа данных в социологии	Ростов-на-Дону, Таганрог: Издательство Южного федерального университета	2023	https://www.iprbookshop.ru/135652.html
Полякова, Э. И.	Экономический и маркетинговый анализ	Москва, Вологда: Инфра-Инженерия	2023	https://www.iprbookshop.ru/133029.html
Кучерова, Е. В., Казарян, М. Т., Королева, Т. Г.	Комплексный экономический анализ финансово-хозяйственной деятельности	Кемерово: Кузбасский государственный технический университет имени Т.Ф. Горбачева	2023	https://www.iprbookshop.ru/135104.html
Вода, И. С., Гракова, М. А., Савельева, Т. М., Андреева, Е. Ю., Палига, Н. Б., Полухина, М. Н.	Современный стратегический анализ	Макеевка: Донбасская национальная академия строительства и архитектуры, ЭБС АСВ	2023	https://www.iprbookshop.ru/135138.html
Афонин, В. В., Никулин, В. В.	Анализ и моделирование типовых систем массового обслуживания	Москва, Вологда: Инфра-Инженерия	2023	https://www.iprbookshop.ru/132853.html
Богомолова, Е. В.	Экономический анализ	Липецк: Липецкий государственный технический университет, ЭБС АСВ	2023	https://www.iprbookshop.ru/133449.html
Воронова, Н. С., Бежанов, С. Г., Воронов, С. А., Хангулян, Е. В., Цупко, О. Ю., Романов, А. И.	Анализ и представление результатов эксперимента	Москва: Национальный исследовательский ядерный университет «МИФИ»	2023	https://www.iprbookshop.ru/132675.html
Жукова, Л. В.	Интеллектуальный анализ данных на платформе Loginom	Москва: Издательский Дом МИСиС	2023	https://www.iprbookshop.ru/137524.html
Глушань, В. М.	Основы системного анализа. В 2 частях. Ч.1	Ростов-на-Дону, Таганрог: Издательство Южного федерального университета	2022	https://www.iprbookshop.ru/125707.html
7.1.2 Дополнительная учебная литература				
Мельниченко, А. С.	Математическая статистика и анализ данных	Москва: Издательский Дом МИСиС	2018	http://www.iprbookshop.ru/78563.html
Замятин, А. В.	Интеллектуальный анализ данных	Томск: Издательский Дом Томского государственного университета	2020	https://www.iprbookshop.ru/116889.html
Любимцева, О. Л.	Блочное планирование эксперимента и анализ данных	Нижний Новгород: Нижегородский государственный архитектурно-строительный университет, ЭБС АСВ	2018	http://www.iprbookshop.ru/80885.html
Брюс П.	Практическая статистика для специалистов Data Science: Пер. с англ.	Санкт-Петербург: БХВ-Петербург	2018	https://ibooks.ru/reading.php?short=1&productid=358886
Чубукова, И. А.	Data Mining	Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа	2020	https://www.iprbookshop.ru/89404.html

Шнарева, Г. В., Пономарева, Ж. Г.	Анализ данных	Симферополь: Университет экономики и управления	2019	https://www.iprbookshop.ru/89482.html
Истомина, А. П.	Анализ данных качественных исследований	Ставрополь: Северо-Кавказский федеральный университет	2018	https://www.iprbookshop.ru/92674.html
Воронов, В. И., Воронова, Л. И., Усачев, В. А.	Data Mining - технологии обработки больших данных	Москва: Московский технический университет связи и информатики	2018	http://www.iprbookshop.ru/81324.html

7.2 Перечень профессиональных баз данных и информационно-справочных систем

1. https://elibrary.ru - Научная электронная библиотека eLIBRARY.RU (ресурсы открытого доступа)
2. https://www.rsl.ru - Российская Государственная Библиотека (ресурсы открытого доступа)
3. https://link.springer.com - Международная реферативная база данных научных изданий Springerlink (ресурсы открытого доступа)
4. https://zbmath.org - Международная реферативная база данных научных изданий zbMATH (ресурсы открытого доступа)

7.3 Перечень лицензионного и свободно распространяемого программного обеспечения

MicrosoftOfficeProfessional

Microsoft Windows

R

Python

Notepad++

Microsoft Visual Studio Community

Microsoft Visual Studio Code

Microsoft Visual C++ 2010 Express

7.4 Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Аудитория	Оснащение
Лекционная аудитория	Мультимедийное оборудование, специализированная мебель, доска
Компьютерный класс	Мультимедийное оборудование, компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду