

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Горбунов Алексей Александрович

Должность: Заместитель начальника университета по учебной работе

Дата подписания: 27.08.2024 15:56:48

Уникальный программный ключ:

286e49ee1471d400cc1f45579d51ed7bbf0e9cc7

**Федеральное государственное бюджетное образовательное
Учреждение высшего образования
«Санкт-Петербургский университет
Государственной противопожарной службы МЧС России»**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
**СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В АНАЛИЗЕ
ДАННЫХ**

**Магистратура по направлению подготовки
27.04.03 «Системный анализ и управление»
Направленность (профиль) «Системный анализ и управление в
организационно-технических системах»**

Санкт-Петербург

1. Цели и задачи дисциплины «Современные информационные технологии в анализе данных»

Цели освоения дисциплины «Современные информационные технологии в анализе данных»:

- формирование или совершенствование компетенций, направленных на использование сквозных цифровых технологий, методов и технологий анализа данных в интересах решения задач автоматического управления сложными объектами и принятия управленческих решений.

Перечень компетенций, формируемых в процессе изучения дисциплины «Современные информационные технологии в анализе данных»

Компетенции	Содержание
ОПК-7	способен выбирать методы и разрабатывать на их основе алгоритмы и программы для решения задач автоматического управления сложными объектами
ПК-5	способен применять методы математического и системного анализа и теории принятия решений для исследования функциональных задач управления организационно-техническими системами на основе отечественных и мировых тенденций развития методов, управления, информационных и интеллектуальных технологий

Задачи дисциплины «Современные информационные технологии в анализе данных»

- ознакомление с базовыми методами анализа данных и машинного обучения;
- ознакомление с основными инструментальными средствами анализа данных;
- формирование начальных навыков в решении частных задач анализа данных и машинного обучения с применением современных программных средств.

2. Перечень планируемых результатов обучения по дисциплине «Методы машинного обучения», соотнесенных с планируемыми результатами освоения образовательной программы

Индикаторы достижения компетенции	Планируемые результаты обучения по дисциплине
Категория (группа) общепрофессиональных компетенций: использование профессиональных навыков	
Выбирает методы решения задач автоматического управления ОПК–7.1	Знает
	современные методы анализа данных и машинного обучения ОПК-7.2
	Умеет
	применять современные методы анализа данных в интересах решения задач автоматического управления ОПК-7.2
Тип задачи профессиональной деятельности: организационно-управленческий	
Владеет навыками использования информационных и интеллектуальных технологий ПК-5.3	Знает
	основные инструментальные средства анализа данных и машинного обучения ПК-5.3
	Умеет
	применять современные программные средства интеллектуального анализа данных ПК-5.3

3. Место дисциплины «Современные информационные технологии в анализе данных» в структуре ОПОП

Дисциплина «Современные информационные технологии в анализе данных» относится к обязательной части основной профессиональной образовательной программы магистратуры по направлению подготовки 27.04.03 «Системный анализ и управление», направленность (профиль) «Системный анализ и управление в организационно-технических системах».

4. Структура и содержание дисциплины «Современные информационные технологии в анализе данных»

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

4.1 Распределение трудоемкости учебной дисциплины «Современные информационные технологии в анализе данных» по заочной форме обучения и видам работ

Вид учебной работы	Трудоемкость			
	з.е.	час.	по курсам	
			1	2
Общая трудоемкость дисциплины по учебному плану	3	108	36	72
Контактная работа, в том числе:		12	4	8
Аудиторные занятия		12	4	8
Лекции (Л)		6	4	2
Практические занятия (ПЗ)		6		6
Самостоятельная работа (СРС)		96	32	64
Зачет		+		+

4.2 Содержание дисциплины «Современные информационные технологии в анализе данных», структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

для заочной формы обучения

№ п.п.	Наименование разделов и тем	Всего часов	Количество часов по видам занятий			Контроль	Самостоятельная работа, в том числе консультация
			Лекции	Практич. занятия	Лабораторные занятия		
1	Основы анализа данных	18	2				16
2	Задачи предобработки и разведочного анализа данных	18	2				16
	Итого за 1 курс	36	4				32
3	Инструментальные средства анализа данных	72	2	6			64
	Зачет	+				+	
	Итого за 2 курс	72	2	6			64
	Итого по дисциплине	108	6	6			96

4.3 Тематический план для обучающихся

Тема 1.

Лекция. Современные цифровые технологии. Кривая гиперцикла Гартнера. Основные понятия науки о данных. Понятия информация, данные, знания. Интеллектуальный анализ данных. Data mining, Machine Learning. Основные задачи анализа данных. Наборы данных.

Самостоятельная работа. Качество данных, качество решаемых задач.

Определение регрессионной модели, Классификация регрессионных моделей. Ограничения и допущения классического метода наименьших квадратов. Оценка качества регрессионной модели. Проблемы мультиколлинеарности. Аномалии, влиятельные и напряженные наблюдения.

Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Дендрограммы. Метод К-средних. Понятие центроида. Профили кластеров.

Формулировка задачи классификации. Классификационный анализ с обучением. Деревья решений. Алгоритмы построения деревьев решений. Методы и алгоритмы построения деревьев. Алгоритм CART. Определение прекращения построения дерева классификации. Использование нейронных сетей для решения задач классификации. Логистическая регрессия. Сравнение результатов классификации различными методами. Метод k-ближайших соседей.

Оценка качества задач классификации. Таблица сопряженности. Понятие чувствительности и специфичности. ROC-кривая. Ошибки первого и второго рода при решении задач классификации.

Рекомендуемая литература:

основная [1,2];

дополнительная [2].

Тема 2.

Лекция. Типы данных. Качество данных. Разведочный анализ данных. Графические средства анализа данных. Гистограмма, ящичная диаграмма, диаграмма «стебель и листья». Предобработка данных. Пропущенные данные. Дублирование данных. Организация сбора данных. Трансформация данных.

Самостоятельная работа. Квантование. Сэмплинг. Решение задач проверки статистических гипотез в R. Классические методы и понятия статистики в R. Непараметрическая статистика.

Матрица факторных нагрузок. Основное соотношение метода главных компонент. Оценка дисперсии отклика. Критерий «каменистой осыпи».

Рекомендуемая литература:

основная [1,2];

дополнительная [2].

Тема 3.

Лекция. Инструментальные средства анализа данных и машинного обучения. Общая характеристика Rstudio, SPSS, Orange, Deductor, Loginom.

Практические занятия. Примеры решения задач регрессионного анализа в различных программных средах. Использование пакета Deductor для решения задач кластерного анализа. Кластерный анализ в средствах интеллектуального анализа Microsoft Office (на R).

Самостоятельная работа. Понятие повторной выборки, бутстрепа, перестановочного теста, кросс-выборки, складного ножа, механизмы генерации случайных псевдовыборок. Редукция данных. Метод главных компонент. Многомерное шкалирование. Основные положения метода анализа главных компонент. Понятие фактора.

Рекомендуемая литература:

основная [2,3];

дополнительная [1].

5. Методические рекомендации по организации изучения дисциплины «Современные информационные технологии в анализе данных»

При реализации программы дисциплины «Современные информационные технологии в анализе данных» используются лекционные и практические занятия.

Общими целями занятий являются:

- обобщение, систематизация, углубление, закрепление теоретических знаний по конкретным темам дисциплины;
- формирование умений применять полученные знания на практике, реализация единства интеллектуальной и практической деятельности;
- выработка при решении поставленных задач профессионально значимых качеств: самостоятельности, ответственности, точности, творческой инициативы.

Целями лекции являются:

- дать систематизированные научные знания по дисциплине, акцентируя внимание на наиболее сложных вопросах;
- стимулировать активную познавательную деятельность обучающихся, способствовать формированию их творческого мышления.

В ходе практического занятия обеспечивается процесс активного взаимодействия обучающихся с преподавателем; приобретаются практические навыки и умения. Цель практического занятия: углубить и закрепить знания, полученные на лекции, формирование навыков использования знаний для решения практических задач; выполнение тестовых заданий по проверке полученных знаний и умений.

Самостоятельная работа обучающихся направлена на углубление и закрепление знаний, полученных на лекциях и других занятиях, выработку навыков самостоятельного активного приобретения новых, дополнительных знаний, подготовку к предстоящим занятиям.

6. Оценочные материалы по дисциплине «Современные информационные технологии в анализе данных»

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины, проводится в соответствии с содержанием дисциплины по видам занятий в форме опроса и тестирования.

Промежуточная аттестация обеспечивает оценивание промежуточных и окончательных результатов обучения по дисциплине, проводится в форме зачета.

6.1. Примерные оценочные материалы:

6.1.1. Текущего контроля

Типовые вопросы для опроса:

1. Дать определение системы поддержки принятия решений, назвать их состав.
2. Характеризовать хранилища данных.
3. Определить понятие «Анализ данных».
4. Перечислить основные методы и средства анализа данных.
5. Дать определение понятия «машинное обучение».
6. Классифицировать методы повторной выборки.
7. Дать определение бустинга
8. Дать общую характеристику языка R.

Типовые задания для тестирования:

1. 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
2. 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

3. 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

4. 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

5. 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

6. 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

7. 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 развязку 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

8. Виды физической неопределенности данных:

А) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в 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 в 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)

9. При работе по методу К-средних

А) элементы не могут переходить из одного кластера в другой

Б) элементы могут переходить из одного кластера в другой

В) процесс заканчивается при стабилизации кластеров

Г) процесс заканчивается за одну итерацию

10. Если сравнивать Data Mining, машинное обучение и статистику, какая из дисциплин сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов:

А) Data Mining

Б) Машинное обучение

В) Статистика

6.1.2. Промежуточной аттестации

Примерный перечень вопросов, выносимых на зачет

1. Основные понятия науки о данных
2. Понятия информация, данные, знания
3. Интеллектуальный анализ данных. Data mining, Machine Learning.
4. Основные задачи анализа данных
5. Наборы данных. Примеры наборов данных
6. Качество данных, качество решаемых задач
7. Инструментальные средства анализа данных и машинного обучения
8. Общая характеристика Rstudio
9. Общая характеристика SPSS

- 10.Квадрат Гартнера для средств машинного обучения и продвинутой аналитики
- 11.Типы данных. Качество данных
- 12.Разведочный анализ данных
- 13.Графические средства анализа данных
- 14.Гистограмма, ящичная диаграмма, диаграмма «стебель и листья»
- 15.Предобработка данных. Пропущенные данные. Дублирование данных. Организация сбора данных.
- 16.Трансформация данных. Квантование. Сэмплинг
- 17.Решение задач проверки статистических гипотез в R
- 18.Классические методы и понятия статистики в R
- 19.Непараметрическая статистика
- 20.Понятие повторной выборки, бутстрепа, перестановочного теста, кросс-выборки, складного ножа, механизмы генерации случайных псевдовыборок
- 21.Редукция данных
- 22.Метод главных компонент
- 23.Факторный анализ
- 24.Многомерное шкалирование
- 25.Основные положения метода анализа главных компонент

6.2. Шкала оценивания результатов промежуточной аттестации и критерии выставления оценок

Система оценивания включает:

Форма контроля	Показатели оценивания	Критерии выставления оценок	Шкала оценивания
зачет	правильность и полнота ответа	дан правильный, полный ответ на поставленный вопрос, показана совокупность осознанных знаний по дисциплине, доказательно раскрыты основные положения вопросов; могут быть допущены недочеты, исправленные самостоятельно в процессе ответа; дан правильный, недостаточно полный ответ на поставленный вопрос, показано умение выделить существенные и несущественные признаки, причинно-следственные	зачтено

		связи; могут быть допущены недочеты, исправленные с помощью преподавателя; дан недостаточно правильный и полный ответ; логика и последовательность изложения имеют нарушения; в ответе отсутствуют выводы.	
		ответ представляет собой разрозненные знания с существенными ошибками по вопросу; присутствуют фрагментарность, нелогичность изложения; дополнительные и уточняющие вопросы не приводят к коррекции ответа на вопрос.	не зачтено

7. Ресурсное обеспечение дисциплины «Современные информационные технологии в анализе данных»

7.1. Лицензионное и свободно распространяемое программное обеспечение

Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства:

Microsoft Windows 7 Professional – ПО-ВЕ8-834 [Лицензионное]

Microsoft Office Standard 2010 – ПО-413-406 [Лицензионное]

7-Zip – ПО-F33-948 [Свободно распространяемое]

Adobe Acrobat Reader – ПО-F63-948 [Свободно распространяемое]

Google Chrome – ПО-F2C-926 [Свободно распространяемое]

МойОфис Образование – ПО-41В-124 [Свободно распространяемое - Отечественное]

7.2. Профессиональные базы данных и информационные справочные системы

Информационная справочная система – Сервер органов государственной власти Российской Федерации <http://россия.рф/> (свободный доступ); профессиональные базы данных – Портал открытых данных Российской Федерации <https://data.gov.ru/> (свободный доступ); федеральный портал «Российское образование» <http://www.edu.ru> (свободный доступ); система официального опубликования правовых актов в электронном виде <http://publication.pravo.gov.ru/> (свободный доступ); справочная правовая

система «КонсультантПлюс: Студент» [Электронный ресурс]. – Режим доступа: <http://student.consultant.ru/>, (свободный доступ); электронная библиотека университета <http://elib.igps.ru> (авторизованный доступ); электронно-библиотечная система «ЭБС IPR BOOKS» <http://www.iprbookshop.ru> (авторизованный доступ).

7.3. Литература

Основная литература:

1. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/26444.html>
2. Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 308 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/26445.html>
3. Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск : УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053>

Дополнительная литература:

1. Буренин, С. Н. Англоязычный статистический пакет PSPP (свободный аналог SPSS) : учебный практикум / С. Н. Буренин, А. С. Буренина. — Москва : Московский гуманитарный университет, 2017. — 68 с. — ISBN 978-5-906912-85-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS:.. — URL: <https://www.iprbookshop.ru/76588.html>
2. Мельниченко А.С. Математическая статистика и анализ данных : учебное пособие / Мельниченко А.С.. — Москва : Издательский Дом МИСиС, 2018. — 45 с. — ISBN 978-5-906953-62-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/78563.html>

7.4 Материально-техническое обеспечение дисциплины « Современные информационные технологии в анализе данных»

Для проведения и обеспечения занятий используются помещения, которые представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой магистратуры, оснащенные

оборудованием и техническими средствами обучения: автоматизированное рабочее место преподавателя, маркерная доска, мультимедийный проектор, документ-камера, посадочные места обучающихся.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде университета.

Автор: к.т.н., доцент Максимов А.В.