

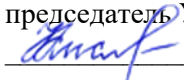
Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Кислова Наталья Николаевна
Должность: Проректор по УМР и качеству образования
Дата подписания: 30.11.2021
Уникальный программный ключ:
52802513f5b14a975b7e9b13008093d5726b159bf6064f865ae65b96a966c035

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Самарский государственный социально-педагогический университет»**

Кафедра информатики, прикладной математики и методики их преподавания

УТВЕРЖДАЮ

Проректор по УМР и КО,
председатель УМС СГСПУ
 Н.Н. Кислова

Введение в анализ данных государственных органов рабочая программа дисциплины (модуля)

Закреплена за кафедрой	Информатики, прикладной математики и методики их преподавания		
Учебный план	ФМФИ-619ПИЗ(4г6м).plx Направление подготовки 09.03.03 Прикладная информатика Направленность (профиль): «Прикладная информатика в государственном и муниципальном управлении» протокол №8 от 29.04.2020 протокол №10 от 26.06.2020		
Квалификация	бакалавр		
Форма обучения	заочная		
Общая трудоемкость	6 ЗЕТ		
Часов по учебному плану	216	Виды контроля в семестрах:	
в том числе:		экзамены 7	
аудиторные занятия	24		
самостоятельная работа	183		
часов на контроль	9		

Распределение часов дисциплины по семестрам

Семестр(Курс.Номер семестра на курсе)	7(4.1)		Итого	
	УП	РПД	УП	РПД
Вид занятий				
Лекции	10	10	10	10
Лабораторные	14	14	14	14
В том числе инт.	6	6	6	6
Итого ауд.	24	24	24	24
Контактная работа	24	24	24	24
Сам. работа	183	183	183	183
Часы на контроль	9	9	9	9
Итого	216	216	216	216

Направление подготовки 09.03.03 Прикладная информатика, направленность (профиль): «Прикладная информатика в государственном и муниципальном управлении»

Рабочая программа дисциплины «Введение в анализ данных государственных органов»

Программу составил(и):

Тюжина Ирина Викторовна

Рабочая программа дисциплины

Введение в анализ данных государственных органов

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.03.03 Прикладная информатика (уровень бакалавриата) (приказ Минобрнауки России от 19.09.2017 г. № 922)

составлена на основании учебного плана:

Направление подготовки 09.03.03 Прикладная информатика

Направленность (профиль): «Прикладная информатика в государственном и муниципальном управлении»

протокол №8 от 29.04.2020

протокол №10 от 26.06.2020

утвержденного учёным советом вуза от 31.08.2018 протокол № 1.

Рабочая программа одобрена на заседании кафедры

Информатики, прикладной математики и методики их преподавания

Протокол от 28.08.2018 г. № 1

Зав. кафедрой Добудько Т.В.

Начальник УОП



Н.А. Доманина

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Цель изучения дисциплины: формирование у студентов универсальной компетенции, связанной со способностью осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач в условиях цифровизации образования.

Задачи изучения дисциплины: сформировать навыки ведения базы данных и поддержки информационного обеспечения решения прикладных задач.

Область профессиональной деятельности: 06 Связь, информационные и коммуникационные технологии

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Цикл (раздел) ОП: Б1.В.08

2.1 Требования к предварительной подготовке обучающегося:

Содержание дисциплины базируется на материале:

Теория вероятностей и математическая статистика

Программирование

Управление данными в корпоративных информационных системах

2.2 Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:

Информационный менеджмент

Управление IT- проектами

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

УК-1.1. Анализирует задачу, выделяя этапы ее решения, действия по решению задачи

Знает: принципы сбора, отбора и обобщения информации, необходимой для анализа открытых данных государственных органов; математические методы анализа открытых данных государственных органов

УК-1.2. Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи

Умеет: анализировать и систематизировать открытые данные государственных органов с применением информационно-коммуникационных технологий; оценивать эффективность различных математических методов анализа данных государственных органов

УК-1.3. Рассматривает различные варианты решения задачи, оценивает их преимущества и риски

Владеет: навыками поиска и практической работы с открытыми данными государственных органов; различными методами решений практических задач дисциплины; методикой оценки результатов решения задач дисциплины.

УК-1.4. Грамотно, логично, аргументированно формирует собственные суждения и оценки; отличает факты от мнений, интерпретаций, оценок в рассуждениях других участников деятельности

Способен аргументировать суждения и оценки, опираясь на статистический анализ открытых данных, предоставляемых органами государственного и муниципального управления

УК-1.5. Определяет и оценивает практические последствия возможных вариантов решения задачи

Способен использовать открытые статистические данные, методы для прогнозирования спроса, продаж программного обеспечения, социально-экономических последствий его внедрения

ПК-5. Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач

ПК-5.1 Знает принципы, технологии и приемы организации баз данных, проектирования архитектуры информационных систем, нормативный и организационные аспекты управления доступа к данным

Знает принципы, технологии и приемы организации баз данных государственных органов, нормативный и организационные аспекты управления доступа к данным

ПК 5.2 Умеет проектировать архитектуру ИС различными инструментальными средствами

Умеет: проектировать компоненты архитектуры ИС, отвечающие за анализ больших данных

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Интеракт.
	Раздел 1.			
1.1	Анализ данных, основные понятия /Лек/	7	2	2
1.2	Введение в машинное обучение. Алгоритмы и структуры данных /Лек/	7	2	
1.3	Фильтрация, группировка и агрегация данных /Лек/	7	2	

Рабочая программа дисциплины «Введение в анализ данных государственных органов»

1.4	Решающие деревья /Лек/	7	2	
1.5	Случайный лес /Лек/	7	2	
1.6	Подготовка и фильтрация данных / Лаб зан /	7	2	
1.7	Преобразование данных / Лаб зан /	7	2	
1.8	Агрегация данных / Лаб зан /	7	2	
1.9	Обучение модели / Лаб зан /	7	2	
1.10	Оценка качества модели /Лаб зан /	7	2	
1.11	Случайный лес / Лаб зан /	7	2	2
1.12	Комплексная лабораторная работа / Лаб зан /	7	4	2
1.13	Анализ источников открытых данных государственных органов /Ср/	7	30	
1.14	Искусственные нейроны /Ср/	7	50	
1.15	Градиентный спуск /Ср/	7	60	
1.16	Визуализация данных /Ср/	7	41	

5. Оценочные и методические материалы по дисциплине (модулю)

5.1. Содержание аудиторной работы по дисциплине (модулю)

Лекция №1

Анализ данных, основные понятия (2 ч)

Данные, модель, обработка данных. Методологические принципы анализа данных. Цели, этапы, методы и техники анализа данных. Средства анализа больших данных. Термин «открытые данные». Доктрина «Открытое государство». Мировой опыт. Механизмы внедрения, цель проекта, комплекс задач, ожидаемый экономический эффект. Законодательная база проекта «Открытые данные».

Лекция №2

Введение в машинное обучение. Алгоритмы и структуры данных (2 ч)

Машинное обучение: базовые понятия и задачи. Python для анализа больших данных. Библиотеки для работы с данными. Физический объём, скорость прироста данных и необходимости их быстрой обработки, возможность одновременно обрабатывать данные различных типов. Библиотека Pandas: структуры данных: Series и DataFrame. Работа с csv.

Лекция №3

Фильтрация, группировка и агрегация данных (2 ч)

Библиотека Pandas. Вывод строк и колонок. Фильтрация данных. Методы query, tail, sample, head. Агрегация данных в Pandas: count, sum, max, min, mean, median. Разница между size и count. Методы groupby и aggregate. Объединение методов.

Лекция №4

Решающие деревья (2 ч)

Классификация. Деревья решений. Основные параметры дерева. Дерево решений в задачи регрессии. Метод ближайших соседей. Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных.

Лекция №5

Алгоритм случайного леса (2 ч)

Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. Тестирование ROC AUC.

Лабораторная работа №1

Подготовка и фильтрация данных (2 ч)

Вопросы и задания:

1. Скачать пакет данных.
2. Избавиться от поврежденных значений.
3. Выбрать данные по заданному признаку.
4. Отобрать данные по группе признаков.

Лабораторная работа №2

Преобразование данных (2 ч)

Вопросы и задания:

1. Формирование нового столбца данных на основе существующих.
2. Формирование вспомогательного столбца.
3. Изменение типа данных столбца.

Лабораторная работа №3

Агрегация данных (2 ч)

Вопросы и задания:

1. Группировка данных по признаку.
2. Агрегация данных: подсчет среднего, общего, максимального и минимального значений по определенному признаку. Выбор оптимального метода.

Лабораторная работа №4
Графическая обработка данных (2 ч)

Вопросы и задания:

1. Построение графиков зависимости.
2. Построение графиков распределения точек, определение числа кластеров, формируемых наблюдениями.
3. Построение тепловой карты.
4. Различные виды графиков (seaborn, matplotlib, встроенная визуализация pandas).

Лабораторная работа №5
Обучение модели (2 ч)

Вопросы и задания:

1. Подготовка данных.
2. Разделение данных на train, test и validation датасеты.
3. Построение дерева решений (sklearn)
4. Визуализация зависимости скоря и аккуратности предсказаний от глубины дерева.

Лабораторная работа №6
Оценка качества модели (2 ч)

Вопросы и задания:

1. Построение confusion matrix.
2. Подсчет Precision, Recall, F1 score.
3. Выбор оптимальных параметров.

Лабораторная работа №7
Случайный лес (2 ч)

Вопросы и задания:

1. Создание модели с указанным числом деревьев и максимальной глубины
2. Обучение модели на данных x_{train} и y_{train} , предскажите класс для наблюдений в x_{test} .
3. Отбор самых топ-5 самых важных переменных для классификации.

Лабораторная работа №8
Комплексная лабораторная работа (4 ч)

Вопросы и задания:

1. Предсказание данных с помощью машинного обучения.
2. Выбор оптимального решения.
3. Оценка качества решения.

5.2. Содержание самостоятельной работы по дисциплине (модулю)

Содержание обязательной самостоятельной работы по дисциплине

№ п/п	Темы дисциплины	Содержание самостоятельной работы студентов	Продукты деятельности
1	Анализ источников открытых данных государственных органов	Поиск данных пригодных для анализа в выбранной области (не менее 5 наборов)	Письменный отчет
2	Искусственные нейроны	Самостоятельно изучить материал, привести пример задачи с решением	Программа
3	Градиентный спуск	Самостоятельно изучить материал, привести пример задачи с решением	Программа
4	Визуализация данных	Самостоятельно изучить материал, привести пример задачи с решением	Программа

Содержание самостоятельной работы по дисциплине на выбор студента

№ п/п	Темы дисциплины	Содержание самостоятельной работы студентов	Продукты деятельности
1	Алгоритм обратного распространения ошибки	Самостоятельно изучить материал, привести пример задачи с решением	Программа

5.3.Образовательные технологии

При организации изучения дисциплины будут использованы следующие образовательные технологии: информационно-коммуникационные технологии, технология организации самостоятельной работы, технология рефлексивного обучения, технология модульного обучения, технология игрового обучения, технологии групповой дискуссии, интерактивные технологии, технология проблемного обучения, технология организации учебно-исследовательской деятельности, технология проектного обучения, технология развития критического мышления.

5.4. Текущий контроль, промежуточный контроль и промежуточная аттестация

Балльно-рейтинговая карта дисциплины оформлена как приложение к рабочей программе дисциплины. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине оформлен отдельным документом.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие, ссылка на электронную библиотечную систему	Издательство, год
Л1.1	Крутиков В.Н., Мешечкин В.В.	Анализ данных: учебное пособие http://biblioclub.ru/index.php?page=book&id=278426	Кемерово: Кемеровский государственный университет, 2014.
Л1.2	Шелудько В.М.	Основы программирования на языке высокого уровня Python: учебное пособие http://biblioclub.ru/index.php?page=book&id=500056	Таганрог: Издательство Южного федерального университета, 2017.

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие, ссылка на электронную библиотечную систему	Издательство, год
Л2.1	Буйначев С.К., Боклаг Н.Ю.	Основы программирования на языке Python: учебное пособие http://biblioclub.ru/index.php?page=book&id=275962	Екатеринбург: Издательство Уральского университета, 2014.

6.2 Перечень программного обеспечения

- АВВУУ Lingvo x6 Многоязычная Академическая версия (30 раб. мест)
- Acrobat Reader DC
- Dr.Web Desktop Security Suite, Dr.Web Server Security Suite
- GIMP
- Microsoft Office 2016 Professional Plus (Пакет программ Word, Excel, Access, PowerPoint, Outlook, OneNote, Publisher)
- Microsoft Office 365 Pro Plus - subscription license (12 month) (Пакет программ Word, Excel, Access, PowerPoint, Outlook, OneNote, Publisher, Skype for Business, OneDrive, SharePoint Online)
- Microsoft Windows 10 Education
- Microsoft Windows 7/8.1 Professional
- RINEL Lingvo v7.0
- XnView
- Архиватор 7-Zip
- НордМастер 5.0, НордКлиент (16 рабочих мест)
- Программная система для обнаружения текстовых заимствований в учебных и научных работах «Антиплагиат.ВУЗ»

6.3 Перечень информационных справочных систем

- Elsevier (база данных «Freedom Collection» и коллекции электронных книг «Freedom Collection eBook collection», национальная подписка на полнотекстовые ресурсы)
- SCOPUS издательства Elsevier
- SpringerNature (национальная подписка на полнотекстовые ресурсы)
- База данных международных индексов научного цитирования Web of Science
- БД «Polpred.com. Обзор СМИ»
- УИС РОССИЯ
- ЭБС «E-LIBRARY.RU»
- ЭБС «ЛАНЬ»
- ЭБС «РУКОНТ» (Контекстум)
- ЭБС «Университетская библиотека онлайн»
- ЭБС «ЮРАЙТ» (Коллекция Легендарные книги)

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Наименование специального помещения: учебная аудитория для проведения занятий лекционного типа, практических занятий, групповых консультаций, индивидуальных консультаций, текущего контроля, промежуточной аттестации, Учебная аудитория. Оснащенность: Меловая доска-1шт., Комплект учебной мебели
7.2	Наименование специального помещения: помещение для самостоятельной работы, Читальный. Оснащенность: ПК -4шт., Письменный стол-4 шт., Парта-2 шт.

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Работа над теоретическим материалом происходит кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю.

Проработка рабочей программы дисциплины, уделяя особое внимание целям и задачам, структуре и содержанию дисциплины. Конспектирование источников, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы, работа с информационными источниками в разных форматах.

Также в процессе изучения дисциплины методические рекомендации могут быть изданы отдельным документом.

Балльно-рейтинговая карта дисциплины «Введение в анализ данных государственных органов»

Курс 4 Семестр 7

Вид контроля		Минимальное количество баллов	Максимальное количество баллов
Наименование раздела			
Текущий контроль по разделу:			
1	Аудиторная работа	40	50
2	Самостоятельная работа (специальные обязательные формы)	5	20
3	Самостоятельная работа (специальные формы на выбор студента)	5	10
Контрольное мероприятие по разделу			-
Промежуточный контроль			6
Итого:		56	100

Виды контроля	Перечень или примеры заданий, критерии оценки и количество баллов	Темы для изучения и образовательные результаты
Текущий контроль по разделу		
1	<p>Аудиторная работа</p> <p>Лабораторная работа (x8) Пример задания Содержание задания: скачайте с портала «Открытые данные России» Данные Информация об обращениях граждан (сведения о тематике и результатах рассмотрения) https://data.gov.ru/opendata/7710349494-og. Подготовьте данные к использованию Средствами языка программирования Python определите наименование вопроса, по которому поступило максимальное количество обращений в Министерство в 4 квартале, и вопросы, по которым поступило минимальное количество обращений. Определите какие вопросы попали в верхние 20% по частоте обращений Критерий оценивания: лабораторные 1-7 - 5 балл – выполнена лабораторная работа, комплексная лабораторная работа – 15 баллов. Итого – 7x5+15=50 баллов</p>	<p>Темы:</p> <p>Подготовка и фильтрация данных Преобразование данных Агрегация данных Графическая обработка данных Обучение модели Оценка качества модели Случайный лес Комплексная лабораторная работа Образовательные результаты: Знает: принципы сбора, отбора и обобщения информации, необходимой для анализа открытых данных государственных органов; математические методы анализа открытых данных государственных органов Умеет: анализировать и систематизировать открытые данные государственных органов с применением информационно-коммуникационных технологий; оценивать эффективность различных математических методов анализа данных государственных органов Владеет: навыками поиска и практической работы с открытыми данными государственных органов; различными методами решений практических задач дисциплины; методикой оценки результатов решения задач дисциплины.</p>

			Способен использовать открытые статистические данные, методы для прогнозирования спроса, продаж программного обеспечения, социально-экономических последствий его внедрения Умеет: проектировать компоненты архитектуры ИС, отвечающие за анализ больших данных
2	Самостоятельная работа (обязательные формы)	<p>Подготовлен текстовый отчет по заданной теме, найдено не менее 5 наборов данных, Критерии оценивания:</p> <ul style="list-style-type: none"> • Отчеты содержат результаты выполнения всех заданий лабораторных работ. • Найденные наборы данных актуальны • Форматы машиночитаемы • Объем данных в наборе соответствует требованиям • Отчет оформлен согласно требованиям и загружен на проверку в систему управления обучением в установленные сроки. <p>Каждый критерий оценивается в 1 балл. Итого – 1x5=5 баллов</p>	<p>Темы: Анализ источников открытых данных государственных органов Искусственные нейроны Градиентный спуск Визуализация данных Образовательные результаты: Знает: принципы сбора, отбора и обобщения информации, необходимой для анализа открытых данных государственных органов; математические методы анализа открытых данных государственных органов Умеет: анализировать и систематизировать открытые данные государственных органов с применением информационно-коммуникационных технологий; оценивать эффективность различных математических методов анализа данных государственных органов Владеет: навыками поиска и практической работы с открытыми данными государственных органов; различными методами решений практических задач дисциплины; методикой оценки результатов решения задач дисциплины.</p>
3	Самостоятельная работа (на выбор студента)	<p>Написана программа по заданной теме Критерии оценивания:</p> <ul style="list-style-type: none"> • Метод анализа выбран корректно • Программа работоспособна • Программа содержит комментарии • На основании работы программы сделаны верные выводы • Отчет оформлен согласно требованиям и загружен на проверку в систему управления обучением в установленные сроки. <p>Каждый критерий оценивается в 2 балла. Итого – 2x5=10 баллов</p>	<p>Темы: Алгоритм обратного распространения ошибки Знает принципы, технологии и приемы организации баз данных государственных органов, нормативный и организационные аспекты управления доступа к данным Умеет: анализировать и систематизировать открытые данные государственных органов с применением информационно-</p>

		коммуникационных технологий; оценивать эффективность различных математических методов анализа данных государственных органов Владеет: навыками поиска и практической работы с открытыми данными государственных органов; различными методами решений практических задач дисциплины; методикой оценки результатов решения задач дисциплины.
Контрольное мероприятие по разделу		-
Промежуточный контроль (количество баллов)	Минимальное количество баллов 6, максимальное - 20	
Промежуточная аттестация	Представлены в фонде оценочных средств для промежуточной аттестации по дисциплине	