

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»



УТВЕРЖДАЮ

Директор НОЦ «Институт
непрерывного образования»

Е.В. Мошкина
Е.В. Мошкина

12 » февраля 2024 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ

«Анализ данных цифрового следа»

Форма обучения – очно-заочная

Срок обучения – 258 часов

Красноярск 2024

УЧЕБНЫЙ ПЛАН
дополнительной профессиональной программы профессиональной переподготовки
«Анализ данных цифрового следа»

Форма обучения – очно-заочная.

Срок обучения – 258 часов.

№ п/п	Наименование модулей (дисциплин)	Общая трудоемкость, ч	Всего контактн., ч	Контактные часы			СРС, ч	Формы контроля
				Лекции	Лабораторные работы	Практические и семинарские занятия		
1.	Математическая статистика	96	64	32		32	32	Экзамен
2.	Анализ данных	108	72	36		36	36	Зачет
3.	Python для сбора и обработки данных цифрового следа	36	18			18	18	Зачет
4.	Итоговая аттестация	18	12			12	6	Защита итоговой аттестационной работы (проекта)
	Итого	258	166	68		98	92	

УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН
дополнительной профессиональной программы профессиональной переподготовки
«Анализ данных цифрового следа»

Категория слушателей: лица, имеющие/получающие высшее образование

Срок обучения: 4 месяца

Форма обучения: очно-заочная (с использованием электронного обучения и дистанционных образовательных технологий)

Режим занятий: до 2-3 часа в день

№ п/п	Наименование модулей (курсов)	Общая трудоемкость, ч	Всего контактн., ч	Контактные часы			СРС, ч	Результаты обучения
				Лекции	Лабораторные работы	Практ. и семинарские занятия		
1	Математическая статистика	96	64	32		32	32	РО 1 – РО 3
1.1	Тема 1.1 Элементы выборочной теории	12	8	4		4	4	РО 1 – РО 3
1.2	Тема 1.2 Оценка параметров распределения	18	12	6		6	6	РО 1 – РО 3
1.3	Тема 1.3. Проверка статистических гипотез	28	22	10		12	6	РО 1 – РО 3
1.4	Тема 1.4. Элементы корреляционного анализа	10	6	4		2	4	РО 1 – РО 3
1.5	Тема 1.5. Регрессионный анализ	14	8	4		4	6	РО 1 – РО 3
1.6	Тема 1.6. Дисперсионный анализ	14	8	4		4	6	РО 1 – РО 3
2	Анализ данных	108	72	36		36	36	РО 4 – РО 6
2.1	Тема 2.1. Описательная статистика и визуальный анализ данных	12	8	4		4	4	РО 4 – РО 6
2.2	Тема 2.2. Алгоритмы обучения с учителем	44	28	14		14	16	РО 4 – РО 6
2.3	Тема 2.3. Алгоритмы обучения без учителя	34	24	12		12	10	РО 4 – РО 6
2.4	Тема 2.4. Анализ связей	18	12	6		6	6	РО 4 – РО 6
3	Python для сбора и обработки данных цифрового следа	36	18			18	18	РО 7 – РО 9
3.1	Тема 3.1 Парсинг и сбор данных цифрового следа	6	4			4	2	РО 7 – РО 9
3.2	Тема 3.2 Структуры данных в Pandas	8	4			4	4	РО 7 – РО 9

№ п/п	Наименование модулей (курсов)	Общая трудоем- кость, ч	Всего контактн., ч	Контактные часы			СРС, ч	Результаты обучения
				Лекции	Лабора- торные работы	Практ. и семинарские занятия		
3.3	Тема 3.3. Визуальный анализ цифрового следа в Matplotlib	8	4			4	4	РО 7 – РО 9
3.4	Тема 3.4. Символьные преобразования в SymPy	8	4			4	4	РО 7 – РО 9
3.5	Тема 3.5. Статистика в SciPy	6	2			2	4	РО 7 – РО 9
	Итоговая аттестация	18	12			12	6	РО 1 – РО 9
	Итого	258	166	68		98	92	

**Календарный учебный график
дополнительной профессиональной программы профессиональной переподготовки
«Анализ данных цифрового следа»**

Наименование модулей (курсов)	Неделя	Объем учебной нагрузки, ч.	Виды занятий (количество часов)							
			Лекция	Практ. и семинарские занятия	Лаб. работа	СРС	Консуль- тация	Контр. работа	Тест	Итоговый контроль
Математическая статистика	1-3	96	32	32		32				Экзамен
Анализ данных	4-7	108	36	36		36				Зачет
Python для сбора и обработки данных цифрового следа	8-10	36		18		18				Зачет
Итоговая аттестация	11-12	18		12		6				Защита итоговой аттестационной работы (проекта)

I. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1. Аннотация программы

Дополнительная профессиональная программа профессиональной переподготовки «Анализ данных цифрового следа» (далее — Программа) ориентирована на подготовку специалистов к новому, перспективному направлению профессиональной деятельности — моделированию, сбору и анализу данных цифрового следа.

Программа предназначена для тех, кто интересуется современными информационными системами и технологиями, в частности технологиями сбора, хранения и анализа больших объемов разнородных данных, программированием, методами искусственного интеллекта.

Программа разработана на основе и с использованием следующих нормативно-правовых и методических документов:

- Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- Методические рекомендации по разработке основных профессиональных образовательных программ и дополнительных профессиональных программ на основе профессиональных стандартов (письмо Минобрнауки РФ от 22 января 2015 г. № ДЛ-1/05);
- Положение о дополнительном образовании и профессиональном обучении в ФГАОУ ВО «Сибирский федеральный университет», утвержденное ректором 01.04.2022 г.;
- Устав ФГАОУ ВО «Сибирский федеральный университет»;
- профессионального стандарта — 06.046 «Специалист по моделированию, сбору и анализу данных цифрового следа», утвержденного Приказом Минтруда России от 09.07.2021, № 462н.

1.2. Цель программы

Цель программы — формирование и совершенствование компетенций, необходимых для выполнения нового вида профессиональной деятельности «Проведение комплексного анализа цифрового следа человека (групп людей) и информационно-коммуникационных систем».

Программа направлена на формирование компетенций в соответствии с трудовыми функциями профессионального стандарта 06.046 «Специалист по моделированию, сбору и анализу данных цифрового следа», утвержденного Приказом Минтруда России от 09.07.2021, № 462н.

Слушатель, успешно завершивший обучение по данной программе, получает диплом о профессиональной переподготовке с правом ведения нового вида профессиональной деятельности в сфере «Проведение комплексного анализа цифрового следа человека (групп людей) и информационно-коммуникационных систем».

Программа является преемственной к основной образовательной программе высшего образования 01.03.04 «Прикладная математика» (уровень

бакалавриата), утвержденной приказом Министерства образования и науки № 208 от 12 марта 2015 г.

1.3. Характеристика нового вида профессиональной деятельности, новой квалификации

1. Область профессиональной деятельности слушателя, прошедшего обучение по Программе, для выполнения нового вида профессиональной деятельности, включает:

- 1) проектирование процесса сбора данных цифрового следа;
- 2) организация работ по подготовке к сбору цифрового следа;
- 3) контроль сбора цифрового следа;
- 4) разработка и применение методов машинного обучения для решения прикладных задач на основе анализа цифрового следа человека (групп людей) и информационно-коммуникационных систем.

2. Объекты профессиональной деятельности:

- 1) открытые источники, а также различные инструменты (рекламные трекеры, файлы Cookies, сервисы по анализу лайков и репостов в социальных сетях);
- 2) данные, относящиеся к высказываниям, профилям в социальных сетях, фотографиям человека;
- 3) учетные записи, подписки на новостные рассылки и публикации, данные о совершенных покупках и использовании приложений, информация о местоположении, настройки браузеров, отзывы и другие данные активного и пассивного цифрового следа;
- 4) методы машинного обучения.

3. Уровень классификации

Программа обеспечивает достижение шестого уровня квалификации в соответствии с требованиями профессионального стандарта 06.046 «Специалист по моделированию, сбору и анализу данных цифрового следа», утвержденного Приказом Минтруда России от 09.07.2021, № 462н.

1.4. Компетенции (трудовые функции) в соответствии с профессиональным стандартом (формирование новых или совершенствование имеющихся)

В соответствии с профессиональным стандартом 06.046 «Специалист по моделированию, сбору и анализу данных цифрового следа», утвержденного Приказом Минтруда России от 09.07.2021, № 462н. можно выделить следующие трудовые функции, на формирование и совершенствование которых направлена Программа:

- С/01.6. Проектирование процесса сбора данных цифрового следа;
- С/02.6. Организация работ по подготовке к сбору цифрового следа;
- С/03.6. Контроль сбора цифрового следа.

1.5. Планируемые результаты обучения

Слушатель, освоивший программу, будет обладать профессиональными компетенциями, включающими в себя способность:

PO1. Применять методы выборочной теории.

PO2. Проводить оценку параметров распределений, проверять статистические гипотезы.

PO3. Применять методы корреляционного, регрессионного, дисперсионного анализа к решению прикладных задач.

PO4. Понимать математические основы моделей прогнозирования на основе данных, поиска закономерностей в данных;

PO5. Подбирать подходящие типы моделей для решения поставленных задач с учетом особенностей исследуемых данных;

PO6. Анализировать достоверность полученных результатов.

PO7. Проектировать процесс сбора данных цифрового следа:

- оценивать необходимое количество данных;

- выбирать метрики оценки качества данных.

PO8. Участвовать в организации работ по сбору данных цифрового следа:

- обезличивать первичные данные человека (групп людей) и ИКС.

PO9. Контролировать сбор цифрового следа:

- контроль соответствия цифрового следа разметке согласно сформулированной модели;

- контроль соответствия процесса получения и обработки данных заданному алгоритму.

1.6. Категория слушателей

Лица, получающие высшее образование по основным образовательным программам бакалавриата, специалитета и магистратуры, отнесенным к укрупнённым группам 01.00.00, 02.00.00, 09.00.00, 10.00.00.

1.7. Требования к уровню подготовки поступающего на обучение

В соответствии с требованиями к образованию и обучению, предъявляемыми к 6 уровню квалификации профессионального стандарта 06.046 «Специалист по моделированию, сбору и анализу данных цифрового следа», необходимо иметь высшее образование или осваивать его в момент обучения на данной программе.

1.8. Продолжительность обучения

Трудоемкость – 258 часов.

1.9. Форма обучения

Очно-заочная (с использованием электронного обучения и дистанционных образовательных технологий).

1.10. Требования к материально-техническому обеспечению, необходимому для реализации дополнительной профессиональной программы профессиональной переподготовки (требования к аудитории, компьютерному классу, программному обеспечению)

Программа повышения квалификации реализуется с использованием системы дистанционного обучения LMS Moodle. Слушателям необходимо стандартное программное обеспечение (операционная система, офисные программы), выход в интернет, а также специализированное программное обеспечение (среда разработки для Python).

1.11. Особенности (принципы) построения дополнительной профессиональной программы профессиональной переподготовки:

- модульная структура программы;
- в основу проектирования программы положен компетентностный подход;
- выполнение комплексных (сквозных) учебных заданий, требующих практического применения знаний и умений, полученных в ходе изучения логически связанных дисциплин (модулей);
- выполнение итоговых аттестационных работ по разработке программного продукта;
- использование информационных и коммуникационных технологий, в том числе современных систем технологической поддержки процесса обучения, обеспечивающих комфортные условия для обучающихся, преподавателей;
- применение электронных образовательных ресурсов (дистанционное, электронное, комбинированное обучение и пр.).

В поддержку дополнительной профессиональной программы профессиональной переподготовки разработаны электронные курсы по всем дисциплинам программы в системе электронного обучения СФУ «e-Курсы» (<https://e.sfu-kras.ru/>).

1.12. Документ об образовании: диплом о переподготовке установленного образца.

II. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

2.1. Формы аттестации, оценочные материалы, методические материалы

Программа предусматривает проведение текущей, промежуточной и итоговой аттестации. Текущая аттестация слушателей проводится на основе оценки активности и участия в дискуссиях в ходе семинаров, а также качества выполнения заданий в электронных обучающих курсах. Промежуточная аттестация осуществляется путем сдачи экзаменов/зачетов по каждой дисциплине.

Методические материалы, необходимые для выполнения текущих заданий, представлены в соответствующих элементах электронных обучающих курсов и включают описание задания, методические рекомендации по его выполнению, критерии оценивания.

Обучение на программе повышения квалификации предполагает выполнение индивидуальных текущих заданий и тестирование.

Итоговой аттестационной работой является дипломный проект.

2.2. Требования и содержание итоговой аттестации

Аттестация проводится в виде зачетов и экзаменов по модулям программы и защиты итоговой аттестационной работы.

К итоговой аттестации допускаются слушатели, выполнившие учебный план программы профессиональной переподготовки, самостоятельные задания в каждой дисциплине в полном объеме за все время обучения.

Итоговая аттестация по программе включает защиту итоговой аттестационной работы (ИАР) в форме проекта. Основная цель итоговой аттестационной работы – выполнить работу, демонстрирующую уровень освоения теоретического и практического материала программы, а также подготовленности к самостоятельной профессиональной деятельности. ИАР выполняется индивидуально.

Слушатель предоставляет результат выполненной работы в формате PDF, оформленной в соответствии с методическими рекомендациями и отвечающей требованиям к содержанию итоговой аттестационной работы.

Защита ИАР проходит в синхронном формате, включает презентацию работы, вопросы по различным разделам программы. Защита ИАР дает возможность продемонстрировать уровень приобретенных слушателем профессиональных компетенций. Объем презентации следует выбирать исходя из длительности выступления (обычно — не более 7–8 минут). В выступлении должны быть четко обозначены тема, область и актуальность работы, постановка цели и задач, приведены результаты, полученные слушателем и проведен их анализ.

Требования к итоговой аттестационной работе

1. Грамотное и качественное выполнение и доработка практических и самостоятельных заданий, выполняемых по мере прохождения программы, которые включены в итоговую аттестационную работу.

2. Полное соблюдение требований к построению текстового документа.
3. Сдача документа в формате pdf с подписями слушателя и научного руководителя на титульном листе.

Критерии оценивания итоговой аттестационной работы

Критерий	Показатели выполнения	Баллы (мин/ макс)
Содержание работы	Обоснована актуальность работы	0/0.5
	Цели и задачи итоговой аттестационной работы определены и согласованы между собой	0/0.5
	Показана практическая значимость работы	0/0.5
	Обоснованы методы анализа данных, применяемые для решения задачи	0/1
	Найдены эффективные варианты решения поставленной задачи	0/4
	Изложение работы иллюстрируется таблицами, диаграммами, схемами, рисунками и т.п.	0/1
	В заключении сформулированы развернутые, самостоятельные выводы	0/0.5
Доклад/ защита работы	Выступление соответствует требованиям публичной речи: материал изложен точно, доступно	0/0.5
	Презентация оформлена в деловом стиле. Информация представлена в виде тезисов, цитат	0/0.5
	Получены ответы на вопросы, заданные членами аттестационной комиссии	0/1
Всего		10 баллов

Оценка «отлично» ставится, если слушатель набрал 9–10 баллов.

Оценка «хорошо» ставится, если слушатель набрал 7–8 баллов.

Оценка «удовлетворительно» ставится, если слушатель набрал 5–6 баллов.

Итоговая аттестационная работа защищается в синхронном формате перед аттестационной комиссией; работа представляется с помощью устного доклада и демонстрации презентации.

Защита итоговой аттестационной работы является обязательной.

Требования к устному докладу в режиме синхронной защиты

1. Приветствие, обращение к членам комиссии и представление (как зовут, кем и где слушатель работает или планирует заниматься), представление научного руководителя.
2. Тема итоговой аттестационной работы.
3. Актуальность, цель и задачи работы.
4. Анализ результатов работы.
5. Заключение.

Продолжительность выступления — 7–8 минут.

По результатам защиты итоговой работы аттестационная комиссия принимает решение о присвоении слушателям по результатам освоения дополнительной профессиональной программы профессиональной переподготовки «Анализ данных цифрового следа» квалификации «Аналитик данных цифрового следа», предоставлении права заниматься профессиональной деятельностью в IT-сфере и выдаче диплома о профессиональной переподготовке установленного образца.

III. ОСНОВНОЕ СОДЕРЖАНИЕ ПРОГРАММЫ

3.1. План учебной деятельности

Результаты обучения	Учебные действия/ формы текущего контроля	Используемые ресурсы/ инструменты/технологии
РО1. Применять методы выборочной теории	Изучение материалов курса, основной и дополнительной литературы, выполнение проектной работы	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО2. Проводить оценку параметров распределений, проверять статистические гипотезы	Изучение материалов курса, основной и дополнительной литературы, выполнение проектной работы	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО3. Применять методы корреляционного, регрессионного, дисперсионного анализа к решению прикладных задач	Изучение материалов курса, основной и дополнительной литературы, выполнение проектной работы	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО4. Понимать математические основы моделей прогнозирования на основе данных, поиска закономерностей в данных	Изучение материалов курса, основной и дополнительной литературы, выполнение проектной работы	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО5. Подбирать подходящие типы моделей для решения поставленных задач с учетом особенностей исследуемых данных	Изучение материалов курса, основной и дополнительной литературы, выполнение проектной работы	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО6. Анализировать достоверность полученных результатов	Изучение материалов курса, основной и дополнительной литературы, выполнение проектной работы	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО7. Проектировать процесс сбора данных цифрового следа	Разбор кейсов, решение мини-задач на программирование. В онлайн-курсе: выполнение тренировочных тестов, выполнение тестов и заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz

Результаты обучения	Учебные действия/ формы текущего контроля	Используемые ресурсы/ инструменты/технологии
РО8. Участвовать в организации работ по сбору данных цифрового следа	Разбор кейсов, решение мини-задач на программирование. В онлайн-курсе: выполнение тренировочных тестов, выполнение тестов и заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz
РО9. Контролировать сбор цифрового следа	Разбор кейсов, решение мини-задач на программирование. В онлайн-курсе: выполнение тренировочных тестов, выполнение тестов и заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции в Sber Jazz

3.2. Виды и содержание самостоятельной работы

Самостоятельная работа слушателя (СРС) предполагает углубление и закрепление теоретических знаний. СРС включает следующие виды самостоятельной деятельности: углубленное изучение вопросов программы, выполнение индивидуальных заданий, подготовка к тестированию и приобретение опыта работы. Выполнение самостоятельной работы слушателями предполагается в дистанционном режиме в рамках электронных курсов, размещенных в системе электронного обучения СФУ «е-Курсы» (<https://e.sfu-kras.ru/>).

РАБОЧАЯ ПРОГРАММА
дисциплины
«Математическая статистика»

1. Аннотация

В рамках данного модуля слушатели ознакомятся с математическим аппаратом, необходимым для анализа случайных явлений и величин; методами обработки экспериментальных данных; приложениями математической статистики к решению прикладных задач.

Цель дисциплины (результаты обучения)

По окончании обучения на данном модуле слушатели будут способны:

РО1. Применять методы выборочной теории.

РО2. Проводить оценку параметров распределений, проверять статистические гипотезы.

РО3. Применять методы корреляционного, регрессионного, дисперсионного анализа к решению прикладных задач.

2. Содержание

№, наименование темы	Содержание лекций (кол-во часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
1. Математическая статистика (96 часов)			
Тема 1.1 Элементы выборочной теории	Основные понятия выборочной теории, 4 часа	Первичная обработка экспериментальных данных, 4 часа	Изучение теоретического материала и решение задач на тему Элементы выборочной теории, 4 часа
Тема 1.2 Оценка параметров распределения	Точечная оценка параметров распределения, 4 часа Интервальная оценка параметров распределения, 2 часа	Решение задач: оценки параметров распределения, 6 часов	Изучение теоретического материала и решение задач на тему Оценка параметров распределения, 6 часов
Тема 1.3. Проверка статистических гипотез	Проверка гипотезы о виде распределения, 6 часов. Проверка гипотез независимости, однородности, случайности, 4 часа.	Решение задач: проверка гипотезы о виде распределения, 6 часов; проверка гипотезы однородности; проверка гипотезы независимости, 6 часов.	Изучение теоретического материала и решение задач на тему Проверка статистических гипотез, 6 часов.

№, наименование темы	Содержание лекций (кол-во часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
Тема 1.4. Элементы корреляционного анализа	Выборочный коэффициент корреляции. Ранговая корреляция, 4 часа	Решение задач: выборочный коэффициент корреляции. Ранговая корреляция, 2 часа	Изучение теоретического материала и решение задач на тему Элементы корреляционного анализа, 4 часа
Тема 1.5. Регрессионный анализ	Регрессионный анализ, 4 часа.	Решение задач: уравнение регрессии, 4 часа.	Изучение теоретического материала и решение задач на тему Регрессионный анализ, 6 часов
Тема 1.6. Дисперсионный анализ	Дисперсионный анализ, 4 часа	Решение задач: однофакторный дисперсионный анализ, 4 часа	Изучение теоретического материала и решение задач на тему Дисперсионный анализ, 6 часов

3. Условия реализации программы модуля

Материально-технические условия реализации программы

Для проведения лекционных занятий необходим мультимедийный проектор. Практические занятия проводятся в компьютерном классе.

Организационно-педагогические условия реализации программы

Обучение по программе реализовано в формате смешанного обучения, с применением активных технологий совместного обучения в электронной среде (синхронные и асинхронные занятия). Лекционный материал представляется в виде комплекса текстовых материалов, размещаемых в LMS Moodle. Данные материалы сопровождаются заданиями. Изучение теоретического материала (СРС) предполагается до и после синхронной части работы.

Учебно-методическое и информационное обеспечение программы модуля

Для выполнения индивидуальных заданий необходима система компьютерной алгебры Mathcad.

Научная электронная библиотека открытого доступа «Киберленинка»
URL: <http://cyberleninka.ru/>

Содержание комплекта учебно-методических материалов

Дисциплина реализуется с применением ЭОК «Математическая статистика», размещенного в системе электронного обучения СФУ «e-Курсы» (<https://e.sfu-kras.ru/>).

Литература

Основная литература

1. Гмурман, В.Е. Теория вероятностей и математическая статистика: учебник для вузов / В.Е. Гмурман. – 12-е изд. – Москва: Издательство Юрайт, 2023. – 479 с.

2. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учебное пособие для вузов / В.Е. Гмурман. - 11-е изд., перераб. и доп. – Москва: Издательство Юрайт, 2023. – 406 с.

Дополнительная литература

1. Крамер, Гаральд. Математические методы статистики = Mathematical Methods of Statistic : перевод с английского / Г. Крамер ; под ред. А. Н. Колмогоров. - 2-е изд., стеротип. - Москва: Мир, 1975. - 648 с.: ил. - 2.87 р. - Текст: непосредственный.

2. Теория вероятностей и математическая статистика. Методы математической статистики и их реализация в среде Mathcad: учеб.-метод. пособие [для студентов напр. 090900.62, 220400.62, 220700.62, 230400.62, 231300.62] / Сиб. федерал. ун-т; сост.: И. И. Вайнштейн, Т. А. Кустицкая. - Электрон. текстовые дан. (PDF, 3,3 Мб). - Красноярск: СФУ, 2012. - Загл. с титул. экрана. - Библиогр.: с. 86. - Текст: электронный.

4. Оценка качества освоения программы модуля (формы аттестации, оценочные и методические материалы)

Форма аттестации по дисциплине — экзамен.

Для допуска к экзамену по дисциплине необходимо:

- получить не менее 40 баллов из 100 за выполнение индивидуальных заданий из ЭОК;

- выполнить тестовые задания к лекциям.

Экзамен проходит в устной форме, включает вопросы по теории и практические задания. По результатам аттестационных испытаний выставляются отметки по системе: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

При осуществлении оценки уровня сформированности компетенций, умений и знаний обучающихся и выставлении отметки целесообразно использовать аддитивный принцип (принцип «сложения»):

отметка «неудовлетворительно» выставляется обучающемуся, не показавшему освоение планируемых результатов (знаний, умений, компетенций), предусмотренных программой, допустившему серьезные ошибки в выполнении предусмотренных программой заданий, не справившемуся с выполнением итоговой аттестационной работы;

отметку «удовлетворительно» заслуживает обучающийся, показавший частичное освоение планируемых результатов (знаний, умений, компетенций), предусмотренных программой, сформированность не в полной мере новых компетенций и профессиональных умений для осуществления профессиональной деятельности, знакомый с литературой, публикациями

по программе. Как правило, отметка «удовлетворительно» выставляется слушателям, допустившим погрешности в итоговой квалификационной работе;

отметку «хорошо» заслуживает обучающийся, показавший освоение планируемых результатов (знаний, умений, компетенций), предусмотренных программой, изучивших литературу, рекомендованную программой, способный к самостоятельному пополнению и обновлению знаний в ходе дальнейшего обучения и профессиональной деятельности;

отметку «отлично» заслуживает обучающийся, показавший полное освоение планируемых результатов (знаний, умений, компетенций), всестороннее и глубокое изучение литературы, публикаций; умение выполнять задания с привнесением собственного видения проблемы, собственного варианта решения практической задачи, проявивший творческие способности в понимании и применении на практике содержания обучения».

Примеры практических заданий

1. Для выборки X найти выборочное среднее и выборочную дисперсию (смещенную и несмещенную).
2. Для выборки X найти размах выборки, моду, медиану.
3. Из равномерного распределения извлечена выборка. С помощью метода моментов оценить параметры распределения.
4. Найти эмпирическую функцию распределения для заданной выборки и построить её график.
5. Дана выборка из совместного распределения СВ X и Y . С помощью метода наименьших квадратов построить уравнение регрессии заданного вида.

Примеры тестовых заданий к лекциям

1. Пусть задана выборка из некоторой генеральной совокупности. С помощью метода моментов можно найти:
 - математическое ожидание и дисперсию соответствующей случайной величины;
 - точечные оценки для параметров распределения соответствующей случайной величины.
 - интервальные оценки для параметров распределения соответствующей случайной величины;
 - эмпирическую функцию распределения.
2. Уровень значимости критерия – это:
 - значение функции от выборки, при которой принимается альтернативная гипотеза;
 - то же, что и уровень доверия;
 - степень важности критерия;
 - вероятность отвергнуть основную гипотезу, когда она верна.
3. Если коэффициент корреляции между X и Y равен нулю, то:
 - между X и Y отсутствует линейная зависимость;
 - X и Y независимы;

- Между X и Y отсутствует нелинейная зависимость;
- X и Y имеют нулевые математические ожидания: $MX=0$, $MY=0$.

4. С помощью метода наименьших квадратов можно построить:

- эмпирическую функцию распределения;
- уравнения нелинейной регрессии;
- точечные оценки параметров распределения;
- интервальные оценки параметров распределения.

5. В критерии Пирсона для проверки гипотезы о виде распределения статистика – это:

- функция от выборки, выражающая отклонение эмпирических частот от теоретических;
- выборка, для которой проверяется гипотеза о виде распределения;
- совокупность интервалов группировки выборочных данных;
- набор проверяемых гипотез.

РАБОЧАЯ ПРОГРАММА

модуля/дисциплины

«Анализ данных»

1. Аннотация

Дисциплина знакомит слушателей с современными методами анализа данных для проведения исследований. В процессе ее изучения слушатели овладеют практическими навыками использования современных программных средств для работы с данными. Они смогут осуществить предварительную подготовку данных для последующей работы с ними в специализированных программах, выбрать подходящий метод анализа в зависимости от типа данных и исследовательской задачи, провести анализ данных и интерпретировать полученные результаты, представлять их в доступном для широкой аудитории виде.

Цель дисциплины (результаты обучения)

Целью изучения дисциплины является ознакомление студентов с основными моделями предиктивного и описательного статистического анализа.

Задачами изучения дисциплины является:

- ознакомление с современными методами анализа статистических данных и сферами их применения в исследовательской практике;
- получение практических навыков применения изученных методов для анализа статистических данных с помощью языка программирования R.

По окончании модуля/дисциплины слушатели будут способны:

РО 4. Понимать математические основы моделей прогнозирования на основе данных, поиска закономерностей в данных;

РО 5. Подбирать подходящие типы моделей для решения поставленных задач с учетом особенностей исследуемых данных;

РО 6. Анализировать достоверность полученных результатов.

2. Содержание

Название темы	Содержание лекций (кол-во часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
Анализ данных (108 часов)			
Тема 2.1. Описательная статистика и визуальный анализ данных (12 ч.)	Машинное обучение и анализ данных (2ч.) Описательная статистика и визуальный анализ данных (2 ч.)	Основы программирования на языке статистических вычислений R (2 ч.) Описательная статистика и визуальный анализ данных (2 ч.)	Описательная статистика и визуальный анализ данных (4 ч.)

Название темы	Содержание лекций (кол-во часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
Анализ данных (108 часов)			
Тема 2.2. Алгоритмы обучения с учителем (44 ч.)	Линейные модели обучения на размеченных данных: регрессия и классификация (4 ч.) Метрики качества (2 ч.) Проблема переобучения (2 ч.) Байесовские модели классификации и регрессии (2 ч.) Метрические алгоритмы (2 ч.) Нейронные сети (2 ч.)	Линейные модели обучения на размеченных данных: регрессия и классификация (4 ч.) Метрики качества (2 ч.) Проблема переобучения (2 ч.) Байесовские модели классификации и регрессии (2 ч.) Метрические алгоритмы (2 ч.) Нейронные сети (2 ч.)	Алгоритмы обучения с учителем (16 ч.)
Тема.2.3. Алгоритмы обучения без учителя (34 ч.)	Задача кластеризации (4 ч.) Метод главных компонент (4 ч.) Выявление ассоциативных правил (4 ч.)	Задача кластеризации (4 ч.) Метод главных компонент (4 ч.) Выявление ассоциативных правил (4 ч.)	Алгоритмы обучения без учителя (10 ч.)
Тема 2.4. Анализ связей (18 ч.)	Критерии различия (3 ч.) Дисперсионный анализ (3 ч.)	Критерии различия (3 ч.) Дисперсионный анализ (3 ч.)	Анализ связей (6 ч.)

3. Условия реализации дисциплины

Материально-технические условия реализации программы

Для проведения лекционных занятий необходим мультимедийный проектор. Практические занятия проводятся в компьютерном классе.

Организационно-педагогические условия реализации программы

Обучение по программе реализовано в формате смешанного обучения, с применением активных технологий совместного обучения в электронной среде (синхронные и асинхронные занятия). Лекционный материал представляется в виде комплекса текстовых материалов/интерактивных ноутбуков, размещаемых в LMS Moodle. Данные материалы сопровождаются заданиями. Изучение теоретического материала (СРС) предполагается до и после синхронной части работы.

Учебно-методическое и информационное обеспечение программы модуля

Для выполнения индивидуальных заданий по программированию алгоритмов анализа данных необходима программная среда вычислений R версии не младше, чем 3.4.3, а также редактор кода RStudio версии 1.1 или старше.

Научная электронная библиотека открытого доступа «Киберленинка»
URL: <http://cyberleninka.ru/>

Содержание комплекта учебно-методических материалов

По данному модулю программы имеется электронный учебно-методический комплекс в LMS Moodle <https://e.sfu-kras.ru/course/view.php?id=22438>. УМК содержит: систему навигации по программе (учебно-тематический план, интерактивный график работы по программе, сведения о результатах обучения, о преподавателях программы, чат для объявлений и вопросов преподавателям), набор интерактивных ноутбук, систему заданий с подробными инструкциями, списки основной и дополнительной литературы.

Литература

Основная литература

1. Ниворожкина Л. И., Арженовский С. В., Рудяга А. А., Торопова Н. А., Федосова О. Н., Житников И. В., Трегубова А. А., Федотова Э. А. Статистические методы анализа данных. Учебник. Москва: Издательский Центр РИО, 2016. – 333 с.
2. Джеймс Г. Введение в статистическое обучение с примерами на языке R. Москва: ДМК Пресс, 2017 1

Дополнительная литература

1. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. Москва: Физматлит: 2006. – 816 с.
2. Кабаков Р., Волкова П. А. R в действии. Анализ и визуализация данных на языке R. Москва: ДМК Пресс, 2014. – 587 с.
3. Волкова П. А., Шипунов А. Б. Статистическая обработка данных в учебно-исследовательских работах. Учебное пособие Москва: Издательство "ФОРУМ": 2016. – 96 с.

4. Оценка качества освоения программы дисциплины

Итоговый контроль проходит в виде зачета по дисциплине.

К зачету допускаются студенты:

- прослушавшие лекции или изучившие их материал самостоятельно с помощью электронного курса и рекомендуемой литературы;
- получили не менее 40 баллов (из 100 возможных) в совокупности за работу на семинарских занятиях, выполнение и защиту индивидуальных заданий, выполнение и защиту группового проекта

Зачет проходит в форме тестирований в электронном курсе.

Тесты на зачет	Содержание заданий	Количество вариантов заданий
1. Тест по теории	Компоненты моделей машинного обучения	4
	Методы решения задач машинного обучения	4
	Линейная регрессия	4
	Метрики качества классификации	4
	Отдельные теоретические вопросы	4
2. Тест по практическим вопросам	Описательный анализ данных	5
	Матрица неточностей	5
	Интерпретация вывода модели	5

Критерии оценивания студента на зачете по дисциплине

Оценка	Требования
«Зачтено»	Студент набрал не менее 40 баллов по каждому из тестов на зачет
«Незачтено»	Студент набрал менее 40 баллов хотя бы по одному из тестов на зачет

Примеры вариантов индивидуальных заданий

Индивидуальное задание.

Линейные методы регрессии

1. Необходимо создать модель прогнозирования цены автомобиля по его характеристикам. Для этого построить 2 модели множественной линейной регрессии

модель 1: в качестве независимых переменных использовать все n числовых признаков (количество числовых признаков определить, проанализировав столбцы своего набора данных).

модель 2: среди всего множества числовых признаков найти 2, наиболее сильно коррелирующие между собой. Тот из признаков, который при этом слабее коррелирует с ценой автомобиля, исключить из рассмотрения. По оставшимся $n-1$ числовым признакам построить модель линейной регрессии для прогнозирования цены авто.

2. Записать (можно на листке бумаги) формулы получившихся регрессий.

3. Проанализировать качество каждой из построенных моделей:

а) найти коэффициент детерминации модели и проинтерпретировать его значение

б) проверить гипотезы о значимости коэффициентов регрессии и уравнения регрессии в целом.

Какая из построенных моделей будет прогнозировать цену авто лучше? Почему?

4. Проверить выполнение условия адекватности каждой из полученных

моделей регрессии (нормальность распределения регрессионных остатков), построив Q-Q график.

5. С помощью обеих моделей найти прогноз цены автомобиля, произвольным образом задав для него значения необходимых числовых признаков

Индивидуальное задание.

Методы классификации

1. Выбрать для проведения классификации один из наборов данных с платформы Kaggle:

нечетные варианты - набор BankMarketing (необходимо построить классификатор, предсказывающий, откроет ли клиент срочный депозит) *файл bank.csv*

четные варианты - набор PimaIndians (необходимо построить классификатор, определяющий, относится ли пациент к группе риска по заболеванию диабетом)

2. Нормировать значения всех количественных признаков.

Категориальные переменные преобразовать в факторные. Разделить выборку на обучающую и контрольную.

3. По обучающей выборке построить 2 классификатора:

- с помощью логистической регрессии (функция **glm()**)
- с помощью метода k ближайших соседей (функция **knn()**), где k
 - номер вашего варианта
 - номер вашего варианта +1

Выведите информацию о параметрах полученных классификаторов

4. Для обеих моделей:

- найти вектор предсказаний классификатора на контрольной выборке
- найти матрицу ошибок классификатора на контрольной выборке
- вычислить метрики качества классификации:
 - accuracy
 - sensitivity
 - specificity
 - precision
 - F-меру

5. Сделать выводы о качестве полученных моделей. С чем могут быть связаны проблемы качества классификации (если они наблюдаются)?

Индивидуальное задание.

Кластеризация

1. Создать набор имитированных данных, получив их с помощью следующей процедуры:

```
x=matrix(rnorm(50*2), ncol=2)
```

```
x[1:25,1]=x[1:25,1]+3
```

```
x[1:25,2]=x[1:25,2]-4
```


2. Выполнить кластеризацию по методу К-средних с $K=2$ (использовать функцию `kmeans()`) 2 раза, используя следующие значения аргумента `nstart`: 1 и 20.

3. Для обеих полученных моделей

– изобразить результаты кластеризации с помощью функции `plot()`, выделяя каждое наблюдение отдельным цветом в соответствии с кластером, к которому оно принадлежит;

– вычислить внутрикластерные суммы квадратов и общую внутрикластерную сумму квадратов (вектора `your_model_name$withinss`, `your_model_name$tot.withinss`), сравнить модели по этим показателям.

4. Аналогично провести кластеризации на 3 кластера.

Какой из видов кластеризации (на 2 или на 3 кластера) выдал наилучший результат и почему?

РАБОЧАЯ ПРОГРАММА
модуля/дисциплины
«Python для сбора и обработки данных цифрового следа»

1. Аннотация

Данный модуль предназначен для знакомства с существующими библиотеками Python. В рамках данного модуля слушатели рассмотрят применение основных пакетов языка Python для анализа данных цифрового следа и решения различных задач.

Цель дисциплины (результаты обучения)

По окончании обучения на данном модуле слушатели будут способны:

РО 7. Проектировать процесс сбора данных цифрового следа:

- оценивать необходимое количество данных;
- выбирать метрики для оценки качества данных.

РО 8. Участвовать в организации работ по сбору данных цифрового следа: обезличивать первичные данные человека (групп людей) и ИКС.

РО 9. Контролировать сбор цифрового следа:

- контроль соответствия цифрового следа разметке согласно сформулированной модели;
- контроль соответствия процесса получения и обработки данных заданному алгоритму.

2. Содержание

№, наименование темы	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
3. Python для сбора и обработки данных цифрового следа (36 часов)		
Тема 3.1 Парсинг и сбор данных цифрового следа (6 ч.)	Парсинг веб-страниц интернет-магазинов (4 ч.)	Изучение теоретических материалов, самостоятельная работа с терминами и определениями (2 ч.)
Тема 3.2 Структуры данных в Pandas (8 ч.)	Цифровой след в Moodle (4 ч.). <i>Задание 1.</i> Извлечение данных цифрового следа Moodle	Сравнение структур данных в Pandas. Изучение структуры логов ЭОК Moodle (4 ч.)
Тема 3.3. Визуальный анализ цифрового следа в Matplotlib (8 ч.)	Построение диаграмм в различных нотациях (4 ч.)	Диаграммы и графики, применяемые для описания данных цифрового следа. Изучение функций библиотеки Matplotlib для отрисовки графиков (4 ч.)
Тема 3.4. Символьные преобразования в SymPy (8 ч.)	Решение задач компьютерной алгебры средствами SymPy (4 ч.)	Знакомство с материалами. Базовые операции и преобразования в SymPy.

№, наименование темы	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
		Сравнение численных и символьных вычислений в Python (4 ч.)
Тема 3.5. Статистика в SciPy (6 ч.)	Основные статистические характеристики выборки в Python (2 ч.) <i>Задание 2.</i> Центральная предельная теорема своими руками	Изучение функций библиотеки SciPy. Статистические распределения в библиотеке SciPy (4 ч.).

3. Условия реализации программы модуля

Материально-технические условия реализации программы

Синхронные занятия реализуются аудиторно и включают в себя лекционные и практические занятия, сочетающие в себе ответы на вопросы, связанные с материалами лекции, в формате дискуссий, а также групповую и индивидуальную работу.

Организационно-педагогические условия реализации программы

Обучение по программе реализовано в формате смешанного обучения, с применением активных технологий совместного обучения в электронной среде (синхронные и асинхронные занятия). Лекционный материал представляется в виде комплекса текстовых материалов/интерактивных ноутбуков, размещаемых в LMS Moodle. Данные материалы сопровождаются заданиями. Изучение теоретического материала (СРС) предполагается до и после синхронной части работы.

Учебно-методическое и информационное обеспечение программы модуля

Программа может быть реализована как очно, так и заочно, в том числе, с применением дистанционных образовательных технологий. Она включает занятия лекционного типа, интерактивные формы обучения, семинарские, мастер-классы, активные и ситуативные методы обучения.

По данному курсу имеется электронный УМК — электронный курс в LMS Moodle.

Содержание комплекта учебно-методических материалов

По данному модулю программы имеется электронный учебно-методический комплекс в LMS Moodle. УМК содержит: систему навигации по программе (учебно-тематический план, интерактивный график работы по программе, сведения о результатах обучения, о преподавателях программы, чат для объявлений и вопросов преподавателям), набор интерактивных ноутбуков, систему заданий с подробными инструкциями, списки основной и дополнительной литературы.

Литература

Основная литература

1. Златопольский Д. М. Основы программирования на языке Python. 2-е изд. – М.: ДМК Пресс, 2018.
2. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения. — Москва: ДМК Пресс, 2017.
3. Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018.

Дополнительная литература

1. Маккинни, Уэс. Python и анализ данных / У. Маккинни. - Электрон. текст. дан. (0.5 MB)col. - [Б. м. : б. и.], печ.20181021030118.0 (ООО "ЛитРес"). - 484 с. : ил ; 479791 байт. - ISBN 978-5-97060-315-4
2. Рамальо, Лучано. Python. К вершинам мастерства / Л. Рамальо. - Электрон. текст. дан. (0.5 MB)col. - [Б. м. : б. и.], печ.20181206235118.0 (ООО "ЛитРес"). - 770 с. : ил ; 538539 байт. - ISBN 978-5-97060-384-0

4. Оценка качества освоения программы модуля (формы аттестации, оценочные и методические материалы)

Форма аттестации по модулю — зачет за выполненные практические задания и за выполненные тесты к лекциям, при условии набора не менее 60 % из 100 верных ответов.

Примеры практических заданий

Задание 1. Извлечение данных цифрового следа Moodle

В данном задании вам необходимо создать таблицу данных в формате .xls/.csv с данными о количестве активных и пассивных действий каждого студента на каждой неделе двух семестров обучения на основе логов цифрового следа.

Таблица по каждому журналу событий должна содержать следующие данные:

1. Фамилия, Имя, Отчество, группа, название электронного курса.
2. Данные о пассивных действиях в курсе для каждой недели `week_N_click`, где N - номер недели обучения.
3. Данные об активных действиях в курсе для каждой недели `week_N_action`, где N - номер недели обучения.

Задание 2. Центральная предельная теорема своими руками

Выберите любимое непрерывное распределение (чем меньше оно будет похоже на нормальное, тем интереснее). Сгенерируйте из него выборку объема 1000, постройте гистограмму выборки и нарисуйте поверх неё теоретическую плотность распределения вашей случайной величины.

Для нескольких значений n

- сгенерируйте 1000 выборок объема n ;
- постройте гистограммы распределений их выборочных средних.

Используя информацию о среднем и дисперсии исходного распределения, посчитайте значения параметров нормальных распределений, которыми, согласно центральной предельной теореме, приближается распределение выборочных средних. Поверх каждой гистограммы нарисуйте плотность соответствующего нормального распределения.

Опишите разницу между полученными распределениями при различных значениях n . Как меняется точность аппроксимации распределения выборочных средних нормальным с ростом n .

Критерии оценивания заданий

Оценка	не зачтено	зачтено
Критерий	Задание выполнено частично, требует серьезной доработки	Задание выполнено, но требует некоторой доработки или Задание выполнено полностью, не требует доработки

Примеры тестовых заданий к лекциям

1. Объект `frame` имеет тип `pandas.DataFrame()`. К чему приведет применение функции `fillna()` к объекту `frame`, в случае вызова функции со следующими параметрами: `frame.fillna("", inplace=True)`?

- Все пропущенные значения в `frame` будут заменены на пустую строку.
- Все пропущенные значения в `frame` будут заменены на 0
- Объект `frame` не будет изменен. В результате запуска будет создана копия объекта `frame`, в которой все пустые строки будут заменены на 0.
- Объект `frame` не будет изменен. В результате запуска будет создана копия объекта `frame`, в которой все пропущенные значения будут заменены на пустую строку.
- Объект `frame` не будет изменен. В результате запуска будет создана копия объекта `frame`, в которой все пропущенные значения будут заменены на 0.

2. Для создания столбчатого графика необходимо использовать следующую функцию:

- `matplotlib.pyplot.barh`
- `matplotlib.pyplot.hist`
- `matplotlib.pyplot.bar`
- `matplotlib.pyplot.boxplot`
- `matplotlib.pyplot.plot`
- `matplotlib.pyplot.pie`

3. Укажите синтаксис специальной функция Бесселя в библиотеке SciPy

- `scipy.special.eppilj()`
- `scipy.special.jn()`
- `scipy.special.gamma()`
- `scipy.special.erf()`

4. Какая библиотека Python предназначена для реализации символьных вычислений?

- a. SymPy
- b. SciPy
- c. Pandas
- d. Scikit-learn
- e. NumPy

5. Объект `x` имеет тип `numpy.ndarray`, какие из перечисленных операций могут при определенных условиях работать с ним?

- a. `x["2, 3"]`
- b. `x[2 3]`
- c. `x[2, 3]`
- d. `x[2][3]`
- e. `x[3:4]`

Программу составили:

Д-р физ.-мат. наук, доцент



И.А. Антипова

Канд. пед. наук



Р.В. Есин

Канд. физ.-мат. наук



А.В. Кошелева

Канд. физ.-мат. наук



И.М. Федотова

Руководитель программы:

Канд. физ.-мат. наук



И.М. Федотова