

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



ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ

«Аналитика данных»

Красноярск 2024

I. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1. Аннотация программы

Дополнительная профессиональная программа профессиональной переподготовки «Аналитика данных» разработана в соответствии с нормами Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»; с учетом требований Постановления Правительства Российской Федерации от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»; с учетом требований приказа Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.02 «Информационные системы и технологии» (уровень бакалавриата), утвержденного приказом Минобрнауки России от 19 сентября 2017 г. № 926, (далее — ФГОС ВО), а также профессионального стандарта 06.042 «Специалист по большим данным», утвержденного приказом Министерства труда и социальной защиты РФ от 6 июля 2020 г. N 405н.

Профессиональная переподготовка заинтересованных лиц (далее — Слушатели), осуществляемая в соответствии с Программой, имеющей отраслевую направленность «Информационно-аналитическое программное обеспечение», проводится в ФГАОУ ВО «Сибирский федеральный университет» (далее — Университет) в соответствии с учебным планом в очно-заочной форме обучения.

Разделы, включенные в учебный план Программы, используются для последующей разработки календарного учебного графика, учебно-тематического плана, рабочих программ модулей (дисциплин), оценочных и методических материалов. Перечисленные документы разрабатываются Университетом самостоятельно, с учетом актуальных положений законодательства об образовании, законодательства в области информационных технологий и смежных областей знаний ФГОС ВО и профессионального стандарта 06.042 «Специалист по большим данным».

Цифровизация широкого круга отраслей деятельности человека привела к тому, что на сегодняшний день генерируются огромные массивы структурированных и неструктурированных данных различной природы, которые необходимо не только хранить, но и извлекать знания, полезные для управления и улучшения качества жизни общества. Кроме этого, для обеспечения технологического суверенитета государства требуется большое число специалистов в области информационных технологий, среди которых немаловажную роль играют специалисты по анализу данных.

1.2. Цель программы

Цель программы профессиональной переподготовки — формирование у слушателей цифровых компетенций в области сбора и подготовки данных для обучения моделей искусственного интеллекта, разработки и применения методов машинного обучения для решения задач, решения задач искусственного интеллекта, а также приобретение по итогам прохождения программы новой квалификации «Аналитик».

Специальность, по которой разработана программа — аналитик данных (Data Scientist).

Слушатель, успешно завершивший обучение по данной программе, получает диплом о профессиональной переподготовке с правом ведения новой профессиональной деятельности в сфере создания и применения технологий больших данных.

1.3. Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и(или) уровней квалификации

1.3.1. Область профессиональной деятельности слушателя, прошедшего обучение по программе профессиональной переподготовки, в которой может осуществлять профессиональную деятельность: разработка и применение методов машинного обучения для решения прикладных задач.

Выпускники могут осуществлять профессиональную деятельность в других областях и (или) сферах профессиональной деятельности при условии соответствия уровня их образования и полученных компетенций требованиям к квалификации работника.

1.3.2. Объекты профессиональной деятельности: методы машинного обучения.

Виды профессиональной деятельности: создание и применение технологий больших данных.

1.3.3. Уровень квалификации. В соответствии с приказом Министерства труда и социальной защиты Российской Федерации от 6 июля 2020 г. «Об утверждении Профессионального стандарта «Специалист по большим данным», дополнительная профессиональная программа профессиональной переподготовки «Аналитика данных» обеспечивает достижение *шестого* уровня квалификации.

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

Программа направлена на формирование компетенций в соответствии с трудовыми функциями профессионального стандарта 06.042 «Специалист по большим данным», утвержденного Приказом Министерства труда и социальной защиты РФ от 06.06.2020 N 405 (таблица).

Характеристика новой квалификации, связанной с видом профессиональной деятельности и трудовыми функциями в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»

Трудовые действия	Трудовая функция	Обобщенная трудовая функция	Вид профессиональной деятельности
Разработка, обсуждение и утверждение содержания аналитических работ с использованием технологий больших данных	A/02.6 Планирование и организация аналитических работ с использованием технологий больших данных	A Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	Создание и применение технологий больших данных
Определение источников больших данных для анализа, идентификация внешних и внутренних источников данных для проведения аналитических работ	A/03.6 Подготовка данных для проведения аналитических работ по исследованию больших данных		
Получение и фильтрация больших объемов данных из гетерогенных источников			
Извлечение, проверка и очистка больших объемов данных из гетерогенных источников			
Оценка соответствия набора данных предметной области и задачам аналитических работ			
Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ	A/04.6 Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика		
Разработка, проверка, оценка используемых моделей больших данных			
Выбор средств представления результатов аналитики больших данных			

1.4. Планируемые результаты обучения

Слушатели в результате освоения программы профессиональной переподготовки «Аналитика данных» смогут:

PO1. Разрабатывать содержание аналитических работ с использованием технологий больших данных.

PO2. Получать и фильтровать данные из гетерогенных источников.

PO3. Извлекать, проверять и чистить данные из гетерогенных источников.

PO4. Проводить оценку соответствия набора данных предметной области и задачам аналитических работ.

PO5. Выбирать методы и инструментальные средства анализа данных для проведения аналитических работ.

PO6. Разрабатывать, проверять и оценивать используемые модели данных.

PO7. Выбирать средства представления результатов аналитики данных.

1.5. Категория слушателей

Лица, желающие получить новую специальность — аналитик данных (Data Scientist), и работать в сфере создания и применения технологий больших данных.

1.6. Требования к уровню подготовки поступающего на обучение

В соответствии с требованиями к образованию и обучению, предъявляемыми к 6 уровню квалификации профессионального стандарта 06.042 «Специалист по большим данным», необходимо иметь высшее образование (бакалавриат).

Для успешного изучения программы слушатели должны иметь:

- знания в области математического анализа, линейной алгебры и теории вероятностей;
- знания синтаксиса и умение программировать на высокоуровневом языке Python.

1.6.1. Описание вступительного испытания

Вступительное испытание является необходимым условием для зачисления на обучение по программе профессиональной переподготовки и предусматривает прохождение тестирования, состоящего из 20 тестовых вопросов. Пороговым значением для успешного прохождения вступительного испытания является достижение не менее 65 % от общего количества результатов выполнения входного теста.

1.7. Продолжительность обучения

256 часов, из них 140 контактных.

1.8. Форма обучения

Очно-заочная (обучение по программе реализовано в формате смешанного обучения, с применением электронного обучения и дистанционных образовательных технологий).

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

Обучение производится на платформе электронного обучения СФУ «е-Курсы» (<https://e.sfu-kras.ru/>). Используются сервисы вебинаров и видеоконференций.

При проведении лекций, практических занятий, самостоятельной работы слушателей используется следующее оборудование: компьютер с наушниками или аудиокolonками, микрофоном и веб-камерой, высокоскоростное подключение к Интернет (не менее 5 Мбит/с).

Программное обеспечение: браузер Google Chrome, Anaconda, PyCharm Community.

1.10. Особенности (принципы) построения дополнительной профессиональной программы профессиональной переподготовки

Особенности построения программы переподготовки «Аналитика данных»:

- в основу проектирования программы положен компетентностный подход;
- выполнение учебных заданий, требующих практического применения знаний и умений, полученных в ходе изучения логически связанных дисциплин;
- выполнение итоговых аттестационных работ по реальному заданию;
- 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 разработан электронный курс <https://e.sfu-kras.ru/course/view.php?id=35455>.

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

УЧЕБНЫЙ ПЛАН
дополнительной профессиональной программы профессиональной переподготовки
«Аналитика данных»

Форма обучения – очно-заочная, с применением электронного обучения и дистанционных образовательных технологий.
 Срок обучения – 256 часов.

№ п/п	Наименование модулей	Общая трудоемкость, ч.	Всего контактн., ч.	Контактные часы			СРС, ч.	Формы контроля
				Лекции	Лаборат. работы	Практические и семинарские занятия		
1	Математическая статистика	72	52	4	-	48	20	Зачет
2	Машинное обучение	112	56	14	-	42	56	Зачет
3	Глубокое обучение	48	24	6	-	18	24	Зачет
4	Итоговая аттестация	24	8	-	-	8	16	Защита итоговой аттестационной работы (проекта)
	Итого	256	140	24	-	116	116	

УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН
дополнительной профессиональной программы профессиональной переподготовки
«Аналитика данных»

Категория слушателей: лица, имеющие или получающие высшее образование.

Срок обучения: 256 часов.

Форма обучения: очно-заочная, с применением электронного обучения и дистанционных образовательных технологий.

Режим занятий: 6 часов в неделю.

№ п/п	Наименование модулей	Общая трудоемкость, ч.	Всего контактн., ч.	Контактные часы			СРС, ч.	Результаты обучения
				Лекции	Лаборат. работы	Практические и семинарские занятия		
1	Математическая статистика	72	52	4	-	48	20	PO4-PO7
1.1	Методы статистического описания результатов наблюдения	36	26	2	-	24	10	PO7
1.2	Статистическое оценивание параметров и проверка гипотез	36	26	2	-	24	10	PO4-PO6
2	Машинное обучение	112	56	14	-	42	56	PO1-PO7
2.1	Введение в машинное обучение	16	8	2	-	6	8	PO1, PO5
2.2	Предварительная обработка и разведочный анализ данных	16	8	2	-	6	8	PO2-PO4
2.3	Задача классификации	16	8	2	-	6	8	PO5-PO7
2.4	Задача восстановления регрессии	16	8	2	-	6	8	PO5-PO7
2.5	Задача кластеризации	16	8	2	-	6	8	PO5-PO7
2.6	Деревья решений и ансамбли моделей	16	8	2	-	6	8	PO5-PO7
2.7	Рекомендательные системы	16	8	2	-	6	8	PO5-PO7
3	Глубокое обучение	48	24	6	-	18	24	PO5-PO7
3.1	Введение в глубокое обучение	16	8	2	-	6	8	PO5-PO7
3.2	Сверточные нейронные сети	16	8	2	-	6	8	PO5-PO7
3.3	Рекуррентные нейронные сети	16	8	2	-	6	8	PO5-PO7
4	Итоговая аттестация	24	8	-	-	8	16	PO1-PO7
	Итого	256	140	24	-	116	116	

II. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

2.1. Формы аттестации, оценочные материалы, методические материалы

Программа предусматривает проведение текущей и итоговой аттестации. Текущая аттестация слушателей проводится по дисциплинам на основе выполнения заданий в электронном обучающем курсе.

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

2.2. Требования и содержание итоговой аттестации

К итоговой аттестации допускаются слушатели, выполнившие учебный план программы, самостоятельные задания в каждой дисциплине. Итоговая аттестация по программе включает защиту итоговой аттестационной работы (ИАР) в форме проекта, которая проходит в синхронном формате.

Основная цель итоговой аттестационной работы — выполнить работу, демонстрирующую уровень подготовленности к самостоятельной профессиональной деятельности.

ИАР выполняется индивидуально. Защита ИАР включает презентацию работы, вопросы по различным разделам программы. Защита ИАР дает возможность продемонстрировать уровень приобретенных слушателем профессиональных компетенций.

Слушатель предоставляет результат выполненной работы в формате PDF, оформленной и отвечающей требованиям к содержанию итоговой аттестационной работы. Список использованных источников литературы приводится в конце ИАР. Документ прикрепляется в организационный электронный курс программы профессиональной переподготовки «Аналитика данных». Объем презентации следует выбирать исходя из длительности выступления (обычно — не более 5–7 минут). В выступлении должны быть четко обозначены область и актуальность работы, постановка задачи, приведены результаты, полученные слушателем. Требования и содержание итоговой аттестации изложены в методических указаниях к выполнению ИАР и размещаются в системе электронного обучения СФУ «e-Курсы».

Примерные темы ИАР

1. Диагностика кожных новообразований.
2. Классификация токсичности комментариев.
3. Распознавание русского жестомимического языка глухонемых.
4. Прогнозирование отказов технологического оборудования.
5. Видеоаналитика производственного травматизма.

Требования к содержанию пояснительной записки ИАР

1. Введение. Описание задачи.
2. Обзорная часть. Описание используемого алгоритма машинного обучения/архитектуры нейронной сети.
3. Постановка задачи машинного обучения.
4. Описание набора данных и шагов по его подготовке для обучения модели.
5. Описание архитектуры модели (моделей) и метода обучения.
6. Оценка качества обучения, графики.
7. Заключение. Описание достигнутых результатов, если использовалось несколько моделей, сравнение качества их работы.
8. Список использованных источников.
9. Приложение. Полный код решения.

Критерии оценивания итоговой аттестационной работы

Оценка «отлично» ставится, если слушатель применил полученные навыки проектирования инструментальных средств анализа данных для решения прикладной задачи системно на экспертном уровне. Контролировал весь цикл программной разработки итогового проекта.

Оценка «хорошо» ставится, если слушатель при выполнении ИАР самостоятельно разработал отдельные модули итогового проекта.

Оценка «удовлетворительно» ставится, если слушатель при выполнении ИАР под контролем преподавателя разработал программное обеспечение.

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.
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 аттестационной работы аттестационная комиссия принимает решение о присвоении слушателям по результатам освоения дополнительной профессиональной программы профессиональной переподготовки квалификации «Аналитик», о предоставлении права заниматься профессиональной деятельностью в сфере создания и применения технологий больших данных и выдаче диплома о профессиональной переподготовке.

III. ОСНОВНОЕ СОДЕРЖАНИЕ ПРОГРАММЫ

3.1. План учебной деятельности

Результаты обучения	Учебные действия/ формы текущего контроля	Используемые ресурсы/ инструменты/технологии
PO1. Разрабатывать содержание аналитических работ с использованием технологий больших данных	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции
PO2. Получать и фильтровать данные из гетерогенных источников	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции
PO3. Извлекать, проверять и чистить данные из гетерогенных источников	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции
PO4. Проводить оценку соответствия набора данных предметной области и задачам аналитических работ	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции
PO5. Выбирать методы и инструментальные средства анализа данных для проведения аналитических работ	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции
PO6. Разрабатывать, проверять и оценивать используемые модели данных	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции
PO7. Выбирать средства представления результатов аналитики данных	Лекции. Выполнение задания, разработка алгоритмов решений в соответствии с указанным заданием. Тесты	Материалы электронного курса в системе электронного обучения СФУ «е-Курсы». Видеоконференции

3.2. Виды и содержание самостоятельной работы

Самостоятельная работа слушателя (СРС) предполагает углубление и закрепление теоретических знаний. СРС включает следующие виды самостоятельной деятельности: самостоятельное углубленное изучение вопросов программы, выполнение индивидуальных заданий, подготовка к тестированию и приобретение опыта работы в рамках электронного курса. Выполнение СРС предполагается в дистанционном режиме в рамках электронного курса.

РАБОЧАЯ ПРОГРАММА
дисциплины (модуля)
«Математическая статистика»

1. Аннотация

В рамках освоения модуля «Математическая статистика» слушатели познакомятся с основными понятиями, задачами и методами математической статистики, научатся проводить первичную обработку выборочных данных, рассчитывать статистические оценки параметров распределений по выборочным данным, проверять статистические гипотезы, овладеют навыками применения методов математической статистики для решения конкретных задач в практических областях.

Цель дисциплины (результаты обучения)

По окончании обучения на данной дисциплине слушатели будут способны:

РО4. Проводить оценку соответствия набора данных предметной области и задачам аналитических работ.

РО5. Выбирать методы и инструментальные средства анализа данных для проведения аналитических работ.

РО6. Разрабатывать, проверять и оценивать используемые модели данных.

РО7. Выбирать средства представления результатов аналитики данных.

2. Содержание

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
Модуль 1. Математическая статистика (72 часа)			
Методы статистического описания результатов наблюдения (36 ч.)	Введение в математическую статистику. Генеральная и выборочная совокупности. Шкалы измерения. Эмпирические функция и плотность распределения. Полигон и гистограмма. Описательные статистики (2 ч.)	Описательный анализ данных на языке Python (24 ч.). <i>Задание 1.</i> Расчет описательных статистик, построение оценок законов распределения данных	Изучение дополнительного материала по темам: Генеральная и выборочная совокупности. Шкалы измерения. Эмпирические функция и плотность распределения. Полигон и гистограмма. Описательные статистики. Тестирование (10 ч.)
Статистическое оценивание параметров и проверка гипотез (36 ч.)	Статистики и их свойства. Оценки статистических характеристик случайных величин.	Проверка статистических гипотез на языке Python (24 ч.). <i>Задание 2.</i>	Изучение дополнительного материала по темам: Статистики и их свойства. Оценки статистических характеристик случайных

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
	Метод максимального правдоподобия. Метод моментов. Доверительная вероятность, доверительный интервал. Общая схема проверки гипотез. Статистический критерий. Критерий согласия. Параметрические и непараметрические подходы проверки гипотез (2 ч.)	Проверка гипотез о виде неизвестного закона распределения генеральной совокупности и неизвестных параметрах известного закона распределения	величин. Метод максимального правдоподобия. Метод моментов. Доверительная вероятность и доверительный интервал. Общая схема проверки гипотез. Статистический критерий. Критерий согласия. Параметрические и непараметрические подходы проверки гипотез. Тестирование (10 ч.)

3. Условия реализации программы дисциплины

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 синхронных лекций, записей занятий, текстовых материалов, презентаций, размещаемых в электронном курсе. Данные материалы сопровождаются заданиями и дискуссиями в чатах дисциплин. Изучение теоретического материала (СРС) предполагается до и после синхронной части работы.

Материально-технические условия реализации программы

Синхронные занятия реализуются на базе инструментов видеоконференцсвязи и включают в себя лекционные и практические занятия. Для проведения синхронных занятий (вебинаров со спикерами) применяется программа видеоконференцсвязи. При проведении лекций, практических занятий, самостоятельной работы слушателей используется следующее оборудование: компьютер с наушниками или аудиоколонками, микрофоном и веб-камерой. Программное обеспечение (обновленное до последней версии): браузер Google Chrome, текстовый редактор.

Учебно-методическое и информационное обеспечение программы

Дисциплина может быть реализована как очно, так и заочно, в том числе, с применением дистанционных образовательных технологий. Она включает занятия лекционного типа, интерактивные формы обучения, практические занятия.

Содержание комплекта учебно-методических материалов

По данной дисциплине имеется электронный учебно-методический комплекс (УМК) в системе электронных курсов СФУ. УМК содержит: систему

навигации по дисциплине (учебно-тематический план, интерактивный график работы по дисциплине, сведения о результатах обучения, чат для объявлений и вопросов преподавателю), текстовые материалы к лекциям, практические и тестовые задания, списки основной и дополнительной литературы. В электронном курсе реализована система обратной связи.

Литература

1. Бондаренко П.С. Теория вероятностей и математическая статистика (для бакалавров) / П.С. Бондаренко, Г.В. Горелова, И.А. Кацко. – М.: КноРус, 2018. – 384 с.
2. Горяинов В.Б. Математическая статистика: учебник для вузов / В.Б. Горяинов, и др.; под ред. В.С. Зарубина, А.П. Крищенко. – Изд. 3-е., испр. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2008. – 423 с.
3. Ивченко Г.И. Математическая статистика / Г.И. Ивченко, Ю.И. Медведев. – М.: КД Либроком, 2019. – 352 с.
4. Кобзарь, А.И. Прикладная математическая статистика. Испр / А.И. Кобзарь. – М.: Физматлит, 2012. – 816 с.
5. Ллойд Э., Ледерман У. Справочник по прикладной статистике. В 2-х томах. – М.: Финансы и статистика, 1990.
6. Пугачёв В.С. Теория вероятностей и математическая статистика: учеб. пособие / В.С. Пугачёв. – М.: Физматлит, 2002. – 496 с.

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

Форма аттестации по дисциплине — зачет.

Оценка результатов обучения осуществляется следующим образом. Максимально за курс можно набрать 100 баллов, из них:

- тесты самоконтроля к лекциям 40 %;
- практические задания составляют 60 %.

Зачет получают слушатели, набравшие не менее 50 % от общего прогресса по курсу.

Примеры тестовых заданий для контроля знаний

1. Оценка называется ..., если ее математическое ожидание равно оцениваемому параметру.
 - а) смещенной;
 - б) несмещенной;
 - в) несостоятельной;
 - г) состоятельной.
2. ... оценкой параметра θ называется числовой интервал, который с заданной точностью покрывает неизвестное значение параметра θ .
 - а) точечной;
 - б) интервальной;
 - в) состоятельной;
 - г) эффективной.

3. Для проверки гипотезы о равенстве дисперсий используется:
- а) критерий Фишера;
 - б) критерий Стьюдента;
 - в) критерий Смирнова;
 - г) критерий Колмогорова.

Типовое практическое задание

Сгенерировать вектор длины 1000 элементов, элементами которого являются реализации нормально распределенной случайной величины с заданными параметрами (математическое ожидание и стандартное отклонение выбрать случайным образом). Подсчитать основные описательные характеристики: минимум и максимум, мат. ожидание, дисперсия, стандартное отклонение, квантили уровней 0.95 и 0.99, медиана, коэффициенты асимметрии, эксцесса. Визуализировать полученные выборки, используя гистограммы. Построить оценки плотностей и визуализировать их. Исследовать отклонение статистического мат. ожидания от заданного при росте объема выборки.

Критерии оценивания заданий

Баллы	3 балла	4 балла	5 баллов
Критерий	Задание выполнено частично, требует серьезной доработки	Задание выполнено, но требует некоторой доработки	Задание выполнено полностью, не требует доработки

РАБОЧАЯ ПРОГРАММА

дисциплины (модуля)

«Машинное обучение»

1. Аннотация

В рамках освоения модуля «Машинное обучение» слушатели познакомятся с основными принципами и методикой анализа данных, основными принципами сбора и подготовки исходных данных, задачей классификации и методами ее решения, основами регрессионного анализа, задачей и методами кластеризации, научатся применять изученные методы анализа данных при решении реальных практических задач, овладеют навыками разработки инструментальных средств анализа данных.

Цель дисциплины (результаты обучения)

По окончании обучения на данной дисциплине слушатели будут способны:

РО1. Разрабатывать содержание аналитических работ с использованием технологий больших данных.

РО2. Получать и фильтровать данные из гетерогенных источников.

РО3. Извлекать, проверять и чистить данные из гетерогенных источников.

РО4. Проводить оценку соответствия набора данных предметной области и задачам аналитических работ.

РО5. Выбирать методы и инструментальные средства анализа данных для проведения аналитических работ.

РО6. Разрабатывать, проверять и оценивать используемые модели данных.

РО7. Выбирать средства представления результатов аналитики данных.

2. Содержание

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
Модуль 2. Машинное обучение (112 часов)			
2.1. Введение в машинное обучение (16 ч.)	Введение в машинное обучение. Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Эффект переобучения и недообучения. Смещение и разброс. Функция потерь и функционал качества.	Введение в машинное обучение на языке Python (6 ч.). <i>Задание 1.</i> Изучение основных инструментов языка программирования Python для анализа данных	Изучение дополнительного материала по темам: Введение в машинное обучение. Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Эффект переобучения и недообучения.

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
	Метод перекрестной проверки. Основные этапы проектов машинного обучения. Особенности накопления выборки данных и количественная его оценка. Разметка данных. Оценка качества работы моделей (2 ч.)		Смещение и разброс. Функция потерь и функционал качества. Метод перекрестной проверки. Основные этапы проектов машинного обучения. Особенности накопления выборки данных и количественная его оценка. Разметка данных. Оценка качества работы моделей. Тестирование (8 ч.)
2.2. Предварительная обработка и разведочный анализ данных (16 ч.)	Ключевые задачи в подготовке данных. Восстановление пропущенных значений. Обнаружение выбросов. Балансировка данных. Обработка категориальных признаков. Масштабирование признаков. Визуализация выборочных данных. Столбиковая диаграмма. Линейный график. Гистограмма. Полигон частот. График плотности распределения вероятностей. График «ящик с усами». Круговая диаграмма. Кольцевая диаграмма. Лепестковая диаграмма. Радиальные диаграммы. Диаграммы рассеяния. Описательная статистика для переменных. Поиск связей между переменными (2 ч.)	Предварительная обработка и разведочный анализ данных на языке Python (6 ч.). <i>Задание 2.</i> Проверка правдоподобности исходных данных. Поиск аномальных значений. Поиск и восстановление пропущенных значений. Преобразование данных. Визуализация данных	Изучение дополнительного материала по темам: Ключевые задачи в подготовке данных. Восстановление пропущенных значений. Обнаружение выбросов. Балансировка данных. Обработка категориальных признаков. Масштабирование признаков. Визуализация выборочных данных. Столбиковая диаграмма. Линейный график. Гистограмма. Полигон частот. График плотности распределения вероятностей. График «ящик с усами». Круговая диаграмма. Кольцевая диаграмма. Лепестковая диаграмма. Радиальные диаграммы. Диаграммы рассеяния. Описательная статистика для переменных. Поиск

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
			связей между переменными. Тестирование (8 ч.)
2.3. Задача классификации (16 ч.)	Постановка задачи классификации. Байесовская теория принятия решений. Логистическая регрессия. Метод ближайших соседей. Метрики оценки качества в задаче классификации (2 ч.)	Классификация данных на языке Python (6 ч.). <i>Задание 3.</i> Обучение моделей классификации. Подбор оптимальных параметров моделей. Оценка качества построенных моделей	Изучение дополнительного материала по темам: Постановка задачи классификации. Байесовская теория принятия решений. Логистическая регрессия. Метод ближайших соседей. Метрики оценки качества в задаче классификации. Тестирование (8 ч.)
2.4. Задача восстановления регрессии (16 ч.)	Постановка задачи восстановления регрессии. Простая линейная регрессия. Множественная линейная регрессия. Метод наименьших квадратов. Метрики оценки качества в задаче восстановления регрессии (2 ч.)	Выполнение регрессионного анализа на языке Python (6 ч.). <i>Задание 4.</i> Построение регрессионных моделей. Подбор оптимальных параметров моделей. Оценка качества построенных моделей	Изучение дополнительного материала по темам: Постановка задачи восстановления регрессии. Простая линейная регрессия. Множественная линейная регрессия. Метод наименьших квадратов. Метрики оценки качества в задаче восстановления регрессии. Тестирование (8 ч.)
2.5. Деревья решений и ансамбли моделей (16 ч.)	Введение в деревья решений. Алгоритмы построения деревьев решений. Редукция решающих деревьев. Ансамблевые методы: Бэггинг, бустинг, стекинг (2 ч.)	Построения деревьев решений и ансамблей алгоритмов на языке Python (6 ч.). <i>Задание 5.</i> Построение моделей на основе деревьев решений и ансамблей алгоритмов (беггинг, бустинг и стекинг) для решения задачи классификации. Подбор оптимальных параметров моделей. Оценка качества	Изучение дополнительного материала по темам: Введение в деревья решений. Алгоритмы построения деревьев решений. Редукция решающих деревьев. Ансамблевые методы: Бэггинг, бустинг, стекинг. Тестирование (8 ч.)

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
		построенных моделей	
2.6. Кластеризация (16 ч.)	Постановка задачи кластеризации. Графовые методы кластеризации. Иерархическая кластеризация. Статистические методы кластеризации (2 ч.)	Выполнение кластерного анализа на языке Python (6 ч.). <i>Задание 6.</i> Построение моделей кластеризации данных. Подбор оптимальных параметров моделей. Оценка качества построенных моделей	Изучение дополнительного материала по темам: Постановка задачи кластеризации. Графовые методы кластеризации. Иерархическая кластеризация. Статистические методы кластеризации. Тестирование (8 ч.)
2.7. Рекомендательные системы (16 ч.)	Коллаборативная фильтрация. Контентная фильтрация. Проблема холодного старта. Метрики оценки качества рекомендательных систем (2 ч.)	Построение рекомендательных систем на языке Python (6 ч.). <i>Задание 7.</i> Создание рекомендательной системы с использованием методов коллаборативной фильтрации и матричной факторизации. Формирование и оценка рекомендаций	Изучение дополнительного материала по темам: Коллаборативная фильтрация. Контентная фильтрация. Проблема холодного старта. Метрики оценки качества рекомендательных систем. Тестирование (8 ч.)

3. Условия реализации программы дисциплины

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 синхронных лекций, записей занятий, текстовых материалов, презентаций, размещаемых в электронном курсе. Данные материалы сопровождаются заданиями и дискуссиями в чатах дисциплин. Изучение теоретического материала (СРС) предполагается до и после синхронной части работы.

Материально-технические условия реализации программы

Синхронные занятия реализуются на базе инструментов видеоконференцсвязи и включают в себя лекционные и практические занятия. Для проведения синхронных занятий (вебинаров со спикерами) применяется программа видеоконференцсвязи. При проведении лекций, практических

занятий, самостоятельной работы слушателей используется следующее оборудование: компьютер с наушниками или аудиокolonками, микрофоном и веб-камерой. Программное обеспечение (обновленное до последней версии): браузер Google Chrome, текстовый редактор.

Учебно-методическое и информационное обеспечение программы

Дисциплина может быть реализована как очно, так и заочно, в том числе, с применением дистанционных образовательных технологий. Она включает занятия лекционного типа, интерактивные формы обучения, практические занятия.

Содержание комплекта учебно-методических материалов

По данной дисциплине имеется электронный учебно-методический комплекс (УМК) в системе электронных курсов СФУ. УМК содержит: систему навигации по дисциплине (учебно-тематический план, интерактивный график работы по дисциплине, сведения о результатах обучения, чат для объявлений и вопросов преподавателю), текстовые материалы к лекциям, практические и тестовые задания, списки основной и дополнительной литературы. В электронном курсе реализована система обратной связи.

Литература

1. Бурков А. Машинное обучение без лишних слов. Библиотека программиста. – М., 2020. – 192 с.
2. Грас Д. Data Science. Наука о данных с нуля: пер. с англ. – СПб.: БХВ-Петербург, 2017. – 336 с.
3. Джоши П. Искусственный интеллект с примерами на Python. – М., 2019. – 448 с.
4. Маккинли У. Python и анализ данных. – Саратов: Профобразование, 2019. – 482 с.
5. Мирджалили В., Рашка С. Python и машинное обучение. Машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow. – М., 2020. – 848 с.
6. Мыльников Л.А. Статистические методы интеллектуального анализа данных. – СПб.: БХВ-Петербург, 2021 – 240 с.
7. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. – М., 2017. – 480 с.
8. Плас Дж. В. Python для сложных задач: наука о данных и машинное обучение. – СПб.: Петербург, 2018 – 576 с.
9. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data, Python и наука о данных. – М., 2017. – 336 с.
10. Скиена С. Наука о данных: учебный курс. Пер. с англ. – СПб.: ООО «Диалектика», 2020. – 544 с.
11. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных: пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2015. – 400 с.

12. Хасти Т., Тибришани Р. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование. – М., 2020. – 768 с.

13. Элбон К. Машинное обучение с использованием Python. Сборник рецептов. – СПб.: БХВ-Петербург, 2020. – 384 с.

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

Форма аттестации по дисциплине — зачет.

Оценка результатов обучения осуществляется следующим образом. Максимально за курс можно набрать 100 баллов, из них:

- тесты самоконтроля к лекциям 40 %;
- практические задания составляют 60 %.

Зачет получают слушатели, набравшие не менее 50 % от общего прогресса по курсу.

Примеры тестов для контроля знаний

1. Множество примеров, используемое для настройки модели, называется:

- а) обучающим множеством;
- б) тестовым множеством;
- в) проверочным множеством.

2. Что является результатом прогноза логистической регрессии?

- а) коэффициенты линейной функции, описывающей закон возникновения события;
- б) численное значение целевой переменной;
- в) погрешность возникновения некоторого события;
- г) вероятность возникновения события.

3. Главной целью кластерного анализа является:

- а) образование групп схожих между собой объектов;
- б) разбиение на группы по некоторым признакам;
- в) различение объектов наблюдения по некоторым признакам;
- г) извлечение наиболее важных признаков из групп данных.

Типовое практическое задание

Для подготовленного набора данных выполнить следующих перечень действий:

1. Провести проверку правдоподобности исходных данных (проверка типов исходных данных, наличие лишних пропусков, невозможных значений и т.п.). Привести найденные значения к нужному формату.
2. Выполнить поиск аномальных значений.
3. Выполнить поиск и восстановление пропущенных значений.
4. Выполнить преобразование данных (привести числовые признаки к стандартному виду).

Критерии оценивания заданий

Баллы	3 балла	4 балла	5 баллов
Критерий	Задание выполнено частично, требует серьезной доработки	Задание выполнено, но требует некоторой доработки	Задание выполнено полностью, не требует доработки

РАБОЧАЯ ПРОГРАММА

дисциплины (модуля)

«Глубокое обучение»

1. Аннотация

В рамках освоения модуля «Глубокое обучение» слушатели познакомятся со способами формирования знаний в области теории нейронных сетей, различными топологиями нейронных сетей, методами их синтеза и способами настройки, научатся применять нейронные сети при решении реальных практических задач, овладеют навыками проектирования нейронных сетей для решения прикладных задач в сфере прогнозирования, распознавания образов, классификации, кластеризации и принятия решений.

Цель дисциплины (результаты обучения)

По окончании обучения на данной дисциплине слушатели будут способны:

РО5. Выбирать методы и инструментальные средства анализа данных для проведения аналитических работ.

РО6. Разрабатывать, проверять и оценивать используемые модели данных.

РО7. Выбирать средства представления результатов аналитики данных.

2. Содержание

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
Модуль 3. Глубокое обучение (48 часов)			
3.1. Основы теории нейронных сетей (16 ч.)	Математический нейрон. Персептрон. Принципы построения нейронных сетей. Однослойные и многослойные нейронные сети. Алгоритмы обучения нейронных сетей. Алгоритм обратного распространения ошибки. Оптимизация в нейронных сетях. Регуляризация в нейронных сетях (2 ч.)	Обучение полносвязной нейронной сети на языке Python (6 ч.). <i>Задание 1.</i> Создание и исследование многослойной нейронной сети для решения задачи классификации. Подбор гиперпараметров нейронной сети	Изучение дополнительного материала по темам: Математический нейрон. Персептрон. Принципы построения нейронных сетей. Однослойные и многослойные нейронные сети. Алгоритмы обучения нейронных сетей. Алгоритм обратного распространения ошибки. Оптимизация в нейронных сетях. Регуляризация в нейронных сетях. Тестирование (8 ч.)

№, наименование темы	Содержание лекций (количество часов)	Наименование практических (семинарских занятий) (кол-во часов)	Виды СРС (кол-во часов)
3.2. Рекуррентные нейронные сети (16 ч.)	Рекуррентная нейронная сеть. Нейронная сеть с долгой краткосрочной памятью (2 ч.)	Обучение рекуррентных нейронных сетей на языке Python (6 ч.). <i>Задание 2.</i> Создание и исследование архитектур рекуррентных нейронных сетей для решения задачи прогнозирования временных рядов. Подбор гиперпараметров нейронной сети	Изучение дополнительного материала по темам: Рекуррентная нейронная сеть. Нейронная сеть с долгой краткосрочной памятью. Тестирование (8 ч.)
3.3. Сверточные нейронные сети (16 ч.)	Сверточные нейронные сети. Архитектуры сетей для классификации изображений. Задачи обработки изображений. Аугментация данных. Transfer learning (2 ч.)	Обучение сверточных нейронных сетей на языке Python (6 ч.). <i>Задание 3.</i> Создание и исследование архитектур сверточных нейронных сетей для решения задачи классификации изображений. Использование трансферного обучения	Изучение дополнительного материала по темам: Сверточные нейронные сети. Архитектуры сетей для классификации изображений. Задачи обработки изображений. Аугментация данных. Transfer learning. Тестирование (8 ч.)

3. Условия реализации программы дисциплины

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 синхронных лекций, записей занятий, текстовых материалов, презентаций, размещаемых в электронном курсе. Данные материалы сопровождаются заданиями и дискуссиями в чатах дисциплин. Изучение теоретического материала (СРС) предполагается до и после синхронной части работы.

Материально-технические условия реализации программы

Синхронные занятия реализуются на базе инструментов видеоконференцсвязи и включают в себя лекционные и практические занятия. Для проведения синхронных занятий (вебинаров со спикерами) применяется программа видеоконференцсвязи. При проведении лекций, практических занятий, самостоятельной работы слушателей используется следующее оборудование: компьютер с наушниками или аудиокolonками, микрофоном и веб-камерой. Программное обеспечение (обновленное до последней версии): браузер Google Chrome, текстовый редактор.

Учебно-методическое и информационное обеспечение программы

Дисциплина может быть реализована как очно, так и заочно, в том числе, с применением дистанционных образовательных технологий. Она включает занятия лекционного типа, интерактивные формы обучения, практические занятия.

Содержание комплекта учебно-методических материалов

По данной дисциплине имеется электронный учебно-методический комплекс (УМК) в системе электронных курсов СФУ. УМК содержит: систему навигации по дисциплине (учебно-тематический план, интерактивный график работы по дисциплине, сведения о результатах обучения, чат для объявлений и вопросов преподавателю), текстовые материалы к лекциям, практические и тестовые задания, списки основной и дополнительной литературы. В электронном курсе реализована система обратной связи.

Литература

1. Будума Н. Основы глубокого обучения. Создание алгоритмов для искусственного интеллекта следующего поколения / Н. Будума, Н. Локашо; пер. с англ. А. Коробейникова; [науч. ред. А. Созыкин]. – М.: Манн, Иванов и Фербер, 2020. – 304 с.

2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А.А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.

3. Келлехер Д. Д. Глубокое обучение. Самый краткий и понятный курс / Д.Д. Келлехер; [перевод с англ. М.А. Райтман]. – М.: Эксмо, 2022. – 160 с.

4. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. – СПб.: Питер, 2018. – 480 с.

5. Паттерсон Дж., Гибсон А. Глубокое обучение с точки зрения практика / пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2018. – 418 с.

6. Траск Э. Грожаем глубокое обучение. – СПб.: Питер, 2019. – 352 с.

7. Шолле Франсуа Глубокое обучение на Python. – СПб.: Питер, 2018. – 400 с.

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

Форма аттестации по дисциплине — зачет.

Оценка результатов обучения осуществляется следующим образом. Максимально за курс можно набрать 100 баллов, из них:

- тесты самоконтроля к лекциям 40 %;
- практические задания составляют 60 %.

Зачет получают слушатели, набравшие не менее 50 % и от общего прогресса по курсу.

Примеры тестов для контроля знаний

1. Нейронная сеть с наличием обратных связей называется:
 - а) полносвязная нейронная сеть;
 - б) обратная нейронная сеть;
 - в) рекуррентная нейронная сеть;
 - г) сверточная нейронная сеть.

2. Гиперпараметрами нейронной сети являются:
 - а) количество нейронов в слое нейронной сети;
 - б) смещение нейронов;
 - в) функции активации;
 - г) функция ошибок;
 - д) веса нейронной сети;
 - е) количество слоев нейронной сети.

3. Механизм, позволяющий определить, какие данные отправить на выход текущей ячейки в LSTM архитектуре
 - а) renewal gate;
 - б) input gate;
 - в) forget gate;
 - г) output gate.

Типовое практическое задание

Для подготовленного набора данных выполнить следующих перечень действий:

1. Построить базовую архитектуру нейронной сети с одним скрытым слоем.
2. Провести тестирование базовой архитектуры нейронной сети.
3. Выполнить подбор гиперпараметров нейронной сети. При подборе гиперпараметров нейронной сети предусмотреть подбор числа слоев, нейронов на каждом слое и функций активации. Выбор архитектуры нейронной сети осуществлять с помощью ошибки на валидационном множестве. После выбора модели на валидационном множестве оценивается её ошибка на тестирующем множестве.
4. Исследовать изменение ошибки обучения для получаемых архитектур нейронных сетей в зависимости от выбранного алгоритма оптимизации.
5. Дополнить архитектуры нейронных сетей блоками для batch-нормализации и дропаута. Сравнить результаты обучения нейронной сети с их использованием и без (скорость сходимости и получаемую точность). Для дропаута подобрать оптимальное значение коэффициента дропаута.

Критерии оценивания заданий

Баллы	3 балла	4 балла	5 баллов
Критерий	Задание выполнено частично, требует серьезной доработки	Задание выполнено, но требует некоторой доработки	Задание выполнено полностью, не требует доработки

Программу составили:

Канд. техн. наук,
доцент базовой кафедры интеллектуальных систем управления Института космических и информационных и технологий СФУ

Е.А. Чжан

Канд. техн. наук,
доцент базовой кафедры интеллектуальных систем управления Института космических и информационных и технологий

А.В. Раскина

Старший преподаватель
научно-учебной лаборатории программного обеспечения Института космических и информационных и технологий СФУ

А.С. Михалев

Старший преподаватель
научно-учебной лаборатории программного обеспечения Института космических и информационных и технологий СФУ

П.В. Пересунько

Руководитель программы:

Канд. техн. наук,
доцент базовой кафедры интеллектуальных систем управления Института космических и информационных и технологий СФУ

А.В. Раскина