

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.05 Методы анализа данных

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

09.03.04 Программная инженерия

Направленность (профиль)

09.03.04 Программная инженерия

Форма обучения

очная

Год набора

2020

Красноярск 2023

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

старший преподаватель, Михалев А.С.

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью преподавания дисциплины «Методы анализа данных» является формирование теоретических знаний по основам методов анализа данных для проведения исследования данных и разработки математических моделей и алгоритмов, выработка практических навыков по применению современных методов решения задач обработки и анализа данных в различных сферах практической деятельности, развитие навыков использования библиотек языка Python для построения программно-алгоритмических решений.

1.2 Задачи изучения дисциплины

В соответствии с требованиями ФГОС ВО к результатам освоения программы бакалавриата задачами изучения дисциплины является:

- изучение соответствующих технологий подготовки данных;
- изучение основных методов обработки знаний и анализа данных;
- формирование умений применения различных методов анализа данных для решения поставленных задач;
- формирование навыков программной реализации методов анализа данных.

Изучение дисциплины «Методы анализа данных» позволяет сформировать у студентов компетенции, необходимые для аналитической, проектной и производственно-технологической деятельности.

В результате изучения курса студент должен:

Знать:

- основные принципы и методику анализа данных;
- основные принципы сбора и подготовки исходных данных;
- основные принципы моделирования неопределённости;
- методы классификации;
- основы регрессионного анализа;
- методы кластеризации.

Уметь:

- применять изученные методы анализа данных при решении реальных практических задач.

Обладать навыками:

- разработки инструментальных средств анализа данных.

1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	
УК-1.1: Осуществляет поиск, анализ информации для решения поставленной задачи	Знать основные методы сбора информации в рамках решаемых задач профессиональной деятельности Уметь выполнять поиск наборов данных в рамках

	<p>решаемой задачи профессиональной деятельности</p> <p>Владеть методами и программными инструментами сбора и анализа информации для решения задач профессиональной деятельности</p>
<p>УК-1.2: Осуществляет критический анализ и синтез информации для решения поставленной задачи</p>	<p>Знать методы предварительной обработки информации</p> <p>Уметь применять инструментальные средства обработки данных при решении задач профессиональной деятельности</p> <p>Владеть навыками формирования сводных массивов статистической информации в соответствии с решаемыми задачами</p>
<p>УК-1.3: Применяет системный подход для решения поставленных задач</p>	<p>Знать современные методы и средства анализа данных, классы решаемых задач</p> <p>Уметь осуществлять выбор наиболее приемлемых методов и подходов к достижению поставленных целей исследования и решения прикладных задач</p> <p>Владеть навыками решения практических задач с помощью методов машинного обучения</p>

1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

URL-адрес и название электронного обучающего курса: <https://e.sfu-kras.ru/course/view.php?id=1185>.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад. час)	е
		1
Контактная работа с преподавателем:	1,5 (54)	
занятия лекционного типа	0,5 (18)	
практические занятия	1 (36)	
Самостоятельная работа обучающихся:	1,5 (54)	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

		Контактная работа, ак. час.							
№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
				Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
		Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
1. Введение в анализ данных									
	1. Введение в анализ данных. Базовые понятия. Общая схема анализа данных. Обучение с учителем, обучение без учителя. Оценка качества обучения	2							
	2. Введение в анализ данных. Базовые понятия. Общая схема анализа данных. Обучение с учителем, обучение без учителя. Оценка качества обучения							2	
2. Первичная обработка данных									
	1. Методы отбора информативных признаков. Регуляризация. Стандартизация и нормализация данных. Восстановление пропущенных значений. Обнаружение выбросов	2							

2. Методы отбора информативных признаков. Регуляризация. Стандартизация и нормализация данных. Восстановление пропущенных значений. Обнаружение выбросов			4					
3. Методы отбора информативных признаков. Регуляризация. Стандартизация и нормализация данных. Восстановление пропущенных значений. Обнаружение выбросов							8	
3. Методы классификации								
1. Задача классификации (постановка и методы решения). Байесовская теория принятия решений. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Непараметрическое оценивание плотности. Логистическая регрессия. Метод ближайших соседей и его обобщения. Стратегии мультиклассовой классификации. Метрики качества классификации		4						
2. Задача классификации (постановка и методы решения). Байесовская теория принятия решений. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Непараметрическое оценивание плотности. Логистическая регрессия. Метод ближайших соседей и его обобщения. Стратегии мультиклассовой классификации. Метрики качества классификации			8					

3. Задача классификации (постановка и методы решения). Байесовская теория принятия решений. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Непараметрическое оценивание плотности. Логистическая регрессия. Метод ближайших соседей и его обобщения. Стратегии мультиклассовой классификации. Метрики качества классификации								10	
4. Задача восстановления регрессии									
1. Задача восстановления регрессии. Простая линейная регрессия. Множественная линейная регрессия. Мультиколлинеарность. Метрики качества регрессии. Нелинейная параметрическая регрессия. Сплайн-функции. Непараметрическая регрессия	4								
2. Задача восстановления регрессии. Простая линейная регрессия. Множественная линейная регрессия. Мультиколлинеарность. Метрики качества регрессии. Нелинейная параметрическая регрессия. Сплайн-функции. Непараметрическая регрессия			8						
3. Задача восстановления регрессии. Простая линейная регрессия. Множественная линейная регрессия. Мультиколлинеарность. Метрики качества регрессии. Нелинейная параметрическая регрессия. Сплайн-функции. Непараметрическая регрессия								10	
5. Деревья решений									
1. Введение в деревья решений. Алгоритмы построения деревьев решений. Редукция решающих деревьев	2								
2. Введение в деревья решений. Алгоритмы построения деревьев решений. Редукция решающих деревьев			6						

3. Введение в деревья решений. Алгоритмы построения деревьев решений. Редукция решающих деревьев							8	
6. Ансамбли моделей								
1. Бэггинг. Бустинг. Стекинг	2							
2. Бэггинг. Бустинг. Стекинг			6					
3. Бэггинг. Бустинг. Стекинг							8	
7. Кластеризация								
1. Введение в кластеризацию. Графовые методы кластеризации. Иерархическая кластеризация. Статистические методы кластеризации	2							
2. Введение в кластеризацию. Графовые методы кластеризации. Иерархическая кластеризация. Статистические методы кластеризации			4					
3. Введение в кластеризацию. Графовые методы кластеризации. Иерархическая кластеризация. Статистические методы кластеризации							8	
Всего	18		36				54	

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Рубан А.И. Методы анализа данных: Учеб.-метод. пособие по курсу «Методы анализа данных» для студентов вузов, обуч. по напр. и спец. информатики и управления: 220100.62, 230102.65, 230105.65 (Красноярск: СФУ).
2. Ту Д. Т., Гонсалес Р. С., Журавлев Ю. И. Принципы распознавания образов: перевод с английского(Москва: Мир).
3. Айвазян С. А., Енюков И. С., Мешалкин Л. Д., Айвазян С. А. Исследование зависимостей: справочное издание(Москва: Финансы и статистика).
4. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д., Айвазян С. А. Прикладная статистика. Классификация и снижение размерности(Москва: Финансы и статистика).
5. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: Часть 1: перевод с английского(Москва: Финансы и статистика).
6. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: Часть 2: перевод с английского(Москва: Финансы и статистика).
7. Лапко А. В., Федотов А. М. Непараметрические методы классификации и их применение: монография(Новосибирск: Наука, Сиб. издат. фирма РАН).
8. Себер Д. А. Ф., Малютков М. Б. Линейный регрессионный анализ: перевод с английского(Москва: Мир).
9. Смит Г. Прикладной регрессионный анализ: Кн. 1: в 2-х кн.(Москва: Финансы и статистика).
10. Смит Г. Прикладной регрессионный анализ: Кн. 2: в 2-х кн.: пер. с англ. (Москва: Финансы и статистика).
11. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д., Айвазян С. А. Прикладная статистика. Классификация и снижение размерности: справ. изд.(Москва: Финансы и статистика).
12. Дуда Р., Харт П., Вайнштейн Г. Г., Васьяковский А. М., Стефанюк В. Л. Распознавание образов и анализ сцен(Москва: Мир).
13. Лапко В. А. Непараметрические методы обработки данных: учебное пособие(Красноярск: ИПЦ КГТУ).
14. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных: справочное издание(М.: Финансы и статистика).
15. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных(Москва: ДМК Пресс).
16. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения(Москва: ДМК Пресс).
17. Вандер Плас Д. Python для сложных задач. Наука о данных и машинное обучение (pdf+epub)(Б. м.: б. и.).

4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):

1. Интегрированная среда разработки программного обеспечения PyCharm.

4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:

1. Электронная библиотечная система «СФУ».

5 Фонд оценочных средств

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Учебные аудитории для проведения лекционных, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, содержащие специализированную мебель, компьютеры с подключением к сети Интернет (неограниченный доступ) и обеспечением доступа к системе виртуальных машин; демонстрационное оборудование (интерактивная доска обратной проекции, проектор, экран для проектора), маркерная доска, доступ к беспроводной сети WI-FI. А также помещение для самостоятельной работы оснащенное компьютерами с подключением к сети Интернет (неограниченный доступ) и обеспечением доступа в электронную информационно-образовательную среду СФУ.