

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

Б1.В.16 Анализ больших данных

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

09.03.02 Информационные системы и технологии

Направленность (профиль)

09.03.02 Информационные системы и технологии

Форма обучения

очная

Год набора

2020

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

Ст. преподаватель, Пахомова Кристина Игоревна;

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью освоения дисциплины «Введение в анализ больших данных» является теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу бакалавриата, для решения различных задач практической и научно-исследовательской деятельности.

1.2 Задачи изучения дисциплины

Задачи освоения дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- приобретение практических навыков работы на языке Python.

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
ПК-3: Способность обеспечения эффективной работы баз данных, включая развертывание, сопровождение, оптимизация функционирования баз данных, являющихся частью различных информационных систем	
ПК-3.1: – знать основные понятия статистики, методы статистических исследований результатов испытаний, основы статистического анализа – знать основные критерии (показатели) работы БД. – знать принципы построения архитектуры программного обеспечения и виды архитектуры программного обеспечения – знать модели и структуры данных, физические модели и архитектуры БД; Особенности реализации структуры данных и управления данными в установленной БД – знать языки и системы программирования БД – знать профессиональные	основные показатели работы БД модели и структуры данных программировать на языке SQL ведениями по работе с БД

<p>сведения о работе БД и методики их применения – знать функционирование компонентов компьютерных систем и принципы их взаимодействия, структуру, компоненты и интерфейсы прикладных систем, взаимодействующих с БД; основы взаимодействия прикладной системы с БД</p>	
---	--

<p>ПК-3.2: – уметь применять автоматизированные средства контроля состояния БД</p> <ul style="list-style-type: none"> – уметь обрабатывать статистические данные, применять методы статистических расчетов – уметь осуществлять самостоятельный поиск информации, необходимой для выполнения профессиональных задач по управлению БД – уметь применять методы оптимизации производительности БД и контролировать полученные результаты – уметь выбирать способ действия в изменяющихся условиях рабочей ситуации; контролировать, оценивать и корректировать свои действия – уметь применять языки и системы программирования БД для оптимизации выполнения запросов – уметь анализировать информацию о работе БД, формулировать выводы, самостоятельно находить информацию, необходимую для описания работы БД, профессионально и понятно письменно излагать выводы, инструкции и рекомендации – уметь локализовать проблемы в БД, понимать причины их возникновения и по возможности, готовить 	
<p>рекомендации по их устранению</p>	

<p>ПК-3.3: – владеть методами мониторинга работы БД, в том числе различными автоматизированными средствами – владеть навыками выбора основных статистических показателей работы БД,</p>	
<p>анализа полученных статистических данных и формирования выводов об эффективности работы БД – владеть навыками анализа возможностей по управлению оптимизацией производительности БД, выбора критериев оптимизации производительности БД; навыками статистического анализа запросов к БД, их классификации по различным признакам, выбора критериев оптимизации запросов к БД, оптимизации выполнения статистически значимых запросов к БД – владеть навыком выработки типовых сценариев восстановления БД при различных сбоях – владеть навыком анализа архитектуры прикладной системы с целью выявления наиболее подверженных сбоям компонентов БД – владеть навыком анализа основных этапов сопровождения БД, подготовки рекомендаций по сопровождению БД, включая оптимизацию критических процессов взаимодействия с БД, подготовки документации в соответствии с установленными правилами и требованиями – владеть навыком сбора информации о проблемах работы пользователей</p>	

прикладной системы с БД – владеть навыком подготовки предложений по выходу из обнаруженных проблемных ситуаций на уровне БД	
--	--

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=21979>.

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	
Промежуточная аттестация (Экзамен)	1 (36)	

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п		Модули, темы (разделы) дисциплины		Контактная работа, ак. час.							
				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 больших данных											
	1. Основные определения, термины, задачи анализа больших данных.	2									
	2. Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data.	2									
	3. Методики сбора данных.	2									
	4. Семантический анализ веб-сайтов			4							
	5. Подготовка к проверочной работе по текущему разделу						10				
2. Технологии хранения и обработки больших данных.											
	1. Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных.	2									
	2. Использование реляционного подхода в представлении данных			2							

3. NoSQL базы данных.	2							
4. Использование NOSQL подхода в представлении данных			2					
5. Подготовка к проверочной работе по текущему разделу							10	
3. Статистические методы анализа данных.								
1. Основные понятия математической статистики. Методы анализа данных.	2							
2. Методы обработки и анализа больших данных.			14					
3. Анализ формальных понятий.	2							
4. Реализация алгоритма АФП			4					
5. Подготовка к проверочной работе по текущему разделу							14	
4. OLAP модель представления данных								
1. Технология OLAP	2							
2. Применение OLAP технологии в практических задачах			4					
3. Подготовка к проверочной работе по текущему разделу							12	
5. Современные программные средства анализа больших объемов информации.								
1. Обзор современных популярных программных средства анализа данных; их преимущества и недостатки.	2							
2. Визуализация исходной информации и аналитических данных.			6					
3. Подготовка к проверочной работе по текущему разделу							8	

Bcero	18		36				54	
-------	----	--	----	--	--	--	----	--

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов(Москва: Высшая школа).
2. Советов Б. Я., Цехановский В. В., Чертовской В. Д. Базы данных: учебник для прикладного бакалавриата(М.: Юрайт).
3. Кроув Т., Эйвисон Д., Вейнеров О. М. Базы данных в административных информационных системах(М.: Финансы и статистика).
4. Маккинли У. Python и анализ данных(Москва: ДМК Пресс).
5. Коэльо Л. П. Построение систем машинного обучения на языке Python (Москва: ДМК Пресс).
6. Бергер А., Горбач И. Microsoft SQL Server 2005 Analysis Services. OLAP и многомерный анализ данных(СПб.: БХВ - Петербург).
7. Сузи Р. А. Python: Пособие(Санкт-Петербург: Издательство "БХВ-Петербург").
8. Дронов В. А. Django. Практика создания Web-сайтов на Python: Пособие (Санкт-Петербург: Издательство "БХВ-Петербург").
9. Прохоренок Н. А. Python 3 и PyQt 5. Разработка приложений: Пособие (Санкт-Петербург: Издательство "БХВ-Петербург").
10. Митчелл Р. Скрапинг веб-сайтов с помощью Python(Москва: ДМК Пресс).
11. Дюк В., Самойленко А. Data Mining. Учебный курс(Санкт-Петербург: Питер).
12. Голицына О. Л., Максимов Н. В., Попов И. И. Базы данных: учебное пособие для учреждений среднего профессионального образования по специальностям "Автоматизированные системы обработки информации и управления (по отраслям)", "Программное обеспечение вычислительной техники и автоматизированных систем", "Техническое обслуживание средств вычислительной техники и компьютерных сетей"(Москва: Форум).
13. Брянцев И.Н. DATA MINING. Теория и практика: к самостоятельной работе(М.: БДЦ - пресс).
14. Прохоренок Н.А. Python. Самое необходимое: Практическое руководство(Санкт-Петербург: Издательство "БХВ-Петербург").
15. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения(Москва: ДМК Пресс).

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. Microsoft Office Professional Plus 2007 Russian OLP NL AE,

2. Python (PSFL Open source) (Python Software Foundation License <https://docs.python.org/3/license.html>),
3. PostgreSQL

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

1. не требуется

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)

Аудитория, оборудованная компьютерами с подключением к сети Интернет и обеспечением доступа в электронную информационно-образовательную среду Университета, доступ к системе виртуальных машин, демонстрационное оборудование:

интерактивная доска обратной проекции; доступ к беспроводной сети WI-FI, маркерная доска.

Занятия организуются с учетом возможности работы обучающихся из числа инвалидов и лиц с ограниченными возможностями в формах, адаптированных к ограничениям их здоровья и восприятия информации в зависимости от нозологии.