

МИНОБРНАУКИ РОССИИ
ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ЭКОНОМИКИ И СЕРВИСА
КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Рабочая программа дисциплины (модуля)
ВВЕДЕНИЕ В АНАЛИЗ БОЛЬШИХ ДАННЫХ

Направление и направленность (профиль)

09.03.02 Информационные системы и технологии. Информационные системы и технологии

Год набора на ОПОП
2018

Форма обучения
заочная

Владивосток 2021

Рабочая программа дисциплины (модуля) «Введение в анализ больших данных» составлена в соответствии с требованиями ФГОС ВО по направлению(ям) подготовки 09.03.02 Информационные системы и технологии (утв. приказом Минобрнауки России от 12.03.2015г. №219) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 05.04.2017 г. N301).

Составитель(и):

Ермолицкая М.З., кандидат биологических наук, доцент, Кафедра информационных технологий и систем, Marina.Ermolitskaya@vvsu.ru

Утверждена на заседании кафедры информационных технологий и систем от 31.05.2021 , протокол № 9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кийкова Е.В.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575633692
Номер транзакции	00000000071ACC1
Владелец	Кийкова Е.В.

Заведующий кафедрой (выпускающей)

Кийкова Е.В.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575633692
Номер транзакции	00000000071ACC2
Владелец	Кийкова Е.В.

1. Цель и задачи освоения дисциплины (модуля)

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.

Задачи освоения дисциплины:

- приобретение студентами знаний о технологиях сбора, хранения и обработки больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- приобретение практических навыков работы с программой R-Studio.

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

Планируемыми результатами обучения по дисциплине являются знания, умения, навыки, соотнесенные с компетенциями, которые формирует дисциплина, и обеспечивающие достижение планируемых результатов по образовательной программе в целом. Перечень компетенций, формируемых в результате изучения дисциплины (модуля), приведен в таблице 1.

Таблица 1 – Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код компетенции	Формулировка компетенции	Планируемые результаты обучения	
09.03.02 «Информационные системы и технологии» (Б-ИС)	ПК-17	Способность использовать технологии разработки объектов профессиональной деятельности в областях: машиностроение, приборостроение, техника, образование, медицина, административное управление, юриспруденция, бизнес, предпринимательство, коммерция, менеджмент, банковские системы, безопасность информационных систем, управление технологическими процессами, механика, техническая физика, энергетика, ядерная энергетика, силовая электроника, металлургия, строительство, транспорт, железнодорожный транспорт, связь, телекоммуникации, управление	Знания:	о современных программных средствах и статистических методах обработки и анализа больших данных
			Умения:	применять теоретические знания для решения задач в области больших данных с использованием современных программных средств

		инфокоммуникациями, почтовая связь, химическая промышленность, сельское хозяйство, текстильная и легкая промышленность, пищевая промышленность, медицинские и би	Навыки:	
				обработки, анализа и представления данных в программе RStudio
	ОПК-5	Способность использовать современные компьютерные технологии поиска информации для решения поставленной задачи, критического анализа этой информации и обоснования принятых идей и подходов к решению	Знания:	технологий сбора, накопления и обработки разнообразной информации
			Умения:	проводить сравнительный анализ и выбор современных информационных технологий поиска и обработки информации
			Навыки:	владения современными информационными технологиями поиска и обработки больших объемов информации

3. Место дисциплины (модуля) в структуре основной образовательной программы

Освоение дисциплины формирует у обучающихся компетенции, необходимые для подготовки бакалавра в соответствии с требованиями ФГОС ВО в области использования современных технологий для сбора, хранения, обработки и анализа больших данных.

Входными требованиями, необходимыми для освоения дисциплины, является наличие у обучающихся компетенций, сформированных при изучении дисциплин и/или прохождении практик «Информатика и основы программирования», «Теория вероятностей и математическая статистика». На данную дисциплину опираются «Защита выпускной квалификационной работы, включая подготовку к процедуре защиты и процедуру защиты», «Производственная преддипломная практика».

4. Объем дисциплины (модуля)

Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обуче- ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо- емкость (З.Е.)	Объем контактной работы (час)					СРС	Форма аттес- тации	
					Всего	Аудиторная			Внеауди- торная			
						лек.	прак.	лаб.	ПА			КСР
09.03.02 Информационные системы и технологии	ЗФО	Бл1.Б	4	3	13	4	8	0	1	0	95	Э

5. Структура и содержание дисциплины (модуля)

5.1 Структура дисциплины (модуля) для ЗФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ЗФО

№	Название темы	Кол-во часов, отведенное на				Форма текущего контроля
		Лек	Практ	Лаб	СРС	
1	Введение в анализ больших данных	4	0	0	50	Собеседование
2	Методы обработки, анализа и визуализации больших данных	0	8	0	45	Проверка практических заданий
Итого по таблице		4	8	0	95	

5.2 Содержание разделов и тем дисциплины (модуля) для ЗФО

Тема 1 Введение в анализ больших данных.

Содержание темы: Основные определения, термины, задачи анализа больших данных. Вопросы безопасности. Понятие Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных. Обзор технологий хранения и представления больших данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Чтение лекций, на которых дается систематизированный материал.

Виды самостоятельной подготовки студентов по теме: Изучение предлагаемой литературы, подготовка к собеседованию.

Тема 2 Методы обработки, анализа и визуализации больших данных.

Содержание темы: Поиск источников информации в сети Интернет: открытые и

закрытые источники данных. Портал открытых данных РФ. Сохранение данных в программе MS Excel. Преобразование и первичная обработка данных. Представление исходных данных в программе RStudio (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах MS Excel и RStudio: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы). Возможности графического представления информации в программе RStudio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Проведение практических занятий в среде R.

Виды самостоятельной подготовки студентов по теме: Изучение предлагаемой литературы.

6. Методические указания по организации изучения дисциплины (модуля)

Программой данной дисциплины предусмотрено чтение лекций, проведение практических занятий и самостоятельная работа студентов для формирования и развития профессиональных навыков обучающихся. На лекционных занятиях с помощью активных и интерактивных методов дается основной систематизированный теоретический материал. На практических занятиях под руководством преподавателя студенты получают навыки сбора, обработки и представления данных, а также приобретают умения применять на практике методы статистического и математического анализа с использованием программы RStudio. Самостоятельная работа студентов заключается в изучении предлагаемой литературы для усвоения и углубления полученных аудиторных знаний. Предусмотренная форма аттестации – экзамен.

Вопросы безопасности больших данных.

Контрольные вопросы для самостоятельной оценки качества освоения учебной дисциплины:

1. Определите сущность понятия «большие данные».
2. Дайте характеристику Big Data на мировом рынке.
3. Охарактеризуйте Big Data в России.
4. Вопросы безопасности больших данных.
5. Технологии хранения больших данных.
6. Источники информации в сети Интернет
7. Методики анализа больших данных.
8. Процесс аналитики анализа больших данных.
9. Определите понятие Data Mining.
10. Когнитивный анализ данных.
11. Основные описательные статистики.
12. Определите различия между параметрическими, непараметрическими и номинальными методами.
13. Основная идея корреляционного анализа.
14. Регрессионный анализ.
15. Основная идея дисперсионного анализа.
16. Сущность кластерного анализа.
17. Дискриминантный анализ: модель и общая процедура выполнения.
18. Цели факторного анализа.
19. Программные средства анализа данных: Statistica, SPSS, Excel; их преимущества и недостатки.
20. Преимущества работа с данными в программе R-Studio.

Для самостоятельной подготовке к занятиям рекомендуется использовать предложенную литературу и информационные ресурсы сети Интернет.

Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов.

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме

электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

7. Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

8. Учебно-методическое и информационное обеспечение дисциплины (модуля)

8.1 Основная литература

1. Гусева Е. Н. Теория вероятностей и математическая статистика : Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Москва : ФЛИНТА , 2016 - 220 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=83543

2. Гутова С. Г., Алтемерова О. А. Теория вероятностей и математическая статистика : Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Кемерово : Кемеровский государственный университет , 2016 - 216 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=481538

3. Кочетков Евгений Семенович. Теория вероятностей и математическая статистика : Учебник [Электронный ресурс] : Форум , 2018 - 240 - Режим доступа: <http://znanium.com/go.php?id=944923>

4. Миркин Б. Г. ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ. Учебник и практикум [Электронный ресурс] : М.:Издательство Юрайт , 2019 - 174 - Режим доступа: <https://biblionline.ru/book/vvedenie-v-analiz-dannyh-432851>

8.2 Дополнительная литература

1. Белько Иван Васильевич. Теория вероятностей, математическая статистика, математическое программирование : Учебное пособие [Электронный ресурс] , 2016 - 299 - Режим доступа: <http://znanium.com/go.php?id=542521>

2. Волкова Полина Андреевна. Статистическая обработка данных в учебно-исследовательских работах : Учебное пособие [Электронный ресурс] : Форум , 2019 - 96 - Режим доступа: <http://znanium.com/go.php?id=1030246>

3. Непомнящая Наталья Васильевна. Статистика: общая теория статистики, экономическая статистика. Практикум : Учебное пособие [Электронный ресурс] , 2015 - 376 - Режим доступа: <http://znanium.com/go.php?id=549841>

4. Основы теории вероятностей и математической статистики [Электронный ресурс] , 2018 - 90 - Режим доступа: <https://lib.rucont.ru/efd/684477>

8.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. - Электронная книга, адрес доступа: <http://r-analytics.blogspot.com>
2. Электронная библиотечная система «РУКОНТ» - Режим доступа: <http://biblioclub.ru/>
3. Электронная библиотечная система «РУКОНТ» - Режим доступа: <https://lib.rucont.ru/>
4. Электронная библиотечная система ZNANIUM.COM - Режим доступа: <http://znanium.com/>
5. Электронная библиотечная система ZNANIUM.COM - Режим доступа: <https://biblio-online.ru/>
6. Open Academic Journals Index (ОАИ). Профессиональная база данных - Режим доступа: <http://oaji.net/>
7. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: <https://www.prlib.ru/>
8. Информационно-справочная система "Консультант Плюс" - Режим доступа: <http://www.consultant.ru/>

9. Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

Основное оборудование:

- Компьютеры
- Монитор облачный 23" LG23CAV42K/мышь Genius Optical Wheel проводная/клавиатура Genius KB110 проводная
- Мультимедийный проектор CASIO (Япония)
- Облачный монитор LG Electronics черный +клавиатура+мышь
- Уст-во бесп.пит.SmartUPS 3000

Программное обеспечение:

- RStudio