

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
"Казанский (Приволжский) федеральный университет"  
Институт вычислительной математики и информационных технологий



*подписано электронно-цифровой подписью*

## Программа дисциплины

Анализ больших данных

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Наука о Данных

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

## Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
  - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
  - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и) научный сотрудник, б/с Заикин А.А. (Научно-исследовательская лаборатория изучения состояния и эволюции подземных резервуаров, Научный центр мирового уровня Рациональное освоение запасов жидких углеводородов планеты (головной центр)), Kaskrin@gmail.com ; доцент, к.н. (доцент) Кареев И.А. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), IAKareev@kpfu.ru ; старший преподаватель, б/с Салимов Р.Ф. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), Rustem.Salimov@kpfu.ru

### 1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-2	Знание и владение методами машинного обучения и анализа данных
ПК-4	Проведение статистического исследования прикладных и теоретических задач

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- содержание действующих российских и международных стандартов в области прикладных программных средств;
- состояние современного рынка прикладных программных продуктов;
- основные подходы к интерпретации и визуализации результатов численных расчетов;
- виды пакетов прикладных программ для использования их в своей профессиональной деятельности.

Должен уметь:

- проводить сбор и первичную обработку данных;
- первичный анализ данных на основе графических возможностей;
- пользоваться аппаратом теории вероятностей и математической статистики для анализа данных;
- пользоваться инструментарием машинного обучения: регрессия, решающие деревья, нейронные сети, кластерный анализ;
- применять полученные знания в прикладных задачах по анализу данных

Должен владеть:

- технологией применения пакетов прикладных программ для решения научных и практических задач.

Должен демонстрировать способность и готовность:

- применять полученные знания в своей учебной и научной деятельности.

### 2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.05 Дисциплины (модули)" основной профессиональной образовательной программы 02.04.02 "Фундаментальная информатика и информационные технологии (Наука о Данных)" и относится к вариативной части.

Осваивается на 2 курсе в 3 семестре.

### 3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 6 зачетных(ые) единиц(ы) на 216 часа(ов).

Контактная работа - 36 часа(ов), в том числе лекции - 18 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 18 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 162 часа(ов).

Контроль (зачёт / экзамен) - 18 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 3 семестре.

#### 4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

##### 4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)			Самостоятельная работа
			Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Определение больших данных. Визуализация больших данных.	3	6	0	6	54
2.	Тема 2. Технологии хранения больших данных. Архитектура решений для анализа больших данных.	3	6	0	6	54
3.	Тема 3. Бутстреп. Анализ данных в реальном времени. Методы понижения размерности.	3	6	0	6	54
	Итого		18	0	18	162

##### 4.2 Содержание дисциплины (модуля)

###### Тема 1. Определение больших данных. Визуализация больших данных.

Лекции:

Определение больших данных. Отличие больших данных от "обычных" данных. Сложности, возникающие при работе с большими данными. Связанные технологии больших данных: облачные вычисления, интернет вещей.

Лабораторные работы:

Работа с пакетами, поддерживающими векторные вычисления.

Самостоятельная работа:

Подбор данных. Изучение методов визуализации больших данных. Выполнение задания на визуализацию данных.

###### Тема 2. Технологии хранения больших данных. Архитектура решений для анализа больших данных.

Лекции:

NoSQL. Различные системы, реализующие принцип NoSQL. Hadoop. Распределённые хранилища. Сбор и транспортировка больших данных.

Лабораторные работы:

Установка и работа с Hadoop. Обработка некоторых запросов для распределенных хранилищ.

Самостоятельная работа:

Настройка и работа с MongoDB. Подготовка бэкенда для распределенного хранилища.

###### Тема 3. Бутстреп. Анализ данных в реальном времени. Методы понижения размерности.

Лекции:

Методы понижения ресурсозатрат при анализе больших данных. Методы, связанные с бутстрепом. Применение методов с различным уровнем сложности. Методы анализа данных в реальном времени. Некоторые нелинейные методы понижения размерности. Отображение на многообразия.

Лабораторные работы:

Применение бутстреп для обучения на больших данных.

Самостоятельная работа:

Реализация одного из методов нелинейного уменьшения размерности и сравнение с линейными методами.

#### 5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства образования и науки Российской Федерации от 5 апреля 2017 года №301)

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

## **6. Фонд оценочных средств по дисциплине (модулю)**

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

## **7. Перечень литературы, необходимой для освоения дисциплины (модуля)**

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы.

Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

## **8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

Habrahabr - [https://habrahabr.ru/hub/machine\\_learning/](https://habrahabr.ru/hub/machine_learning/)

Kaggle - <https://www.kaggle.com>

Школа анализа данных Яндекс - <https://yandexdataschool.ru/edu-process/courses/machine-learning>

## **9. Методические указания для обучающихся по освоению дисциплины (модуля)**

Вид работ	Методические рекомендации
лекции	Во время лекций рекомендуется иметь под рукой (скажем, на электронном носителе) монографию "T. Hastie, R. Tibshirani, J. Friedman. Elements of statistical learning. Data Mining, Inference, and Prediction" поздней редакции (некоторые редакции находятся в открытом доступе). В этой книге содержится множество иллюстраций, которые полезны для понимания материала на лекциях.
лабораторные работы	При выполнении лабораторной работы весьма важна кооперация между студентами при настройке распределенной системы хранения данных. Несмотря на то, что оцениваться будет каждая работа в отдельности, доступ к индивидуальным данным может осуществляться на основе общей системы. Кроме облегчения выполнения задания это поможет студенту лучше понять работу по масштабируемости выбранной системы хранения данных.
самостоятельная работа	В программу курса не входит метод Монте-Карло, поэтому основы этого метода рекомендуется изучить самостоятельно. При выполнении самостоятельных заданий, а также при изучении новых методов, описанных на занятиях, рекомендуется проверять вероятностные утверждения с помощью метода Монте-Карло. Это позволит студенту лучше освоить сам метод, и лучше понять смысл проверяемых утверждений.
экзамен	Следует понимать, что некоторые вопросы могут иметь множество ответов (например, вопрос "На основе чего сравнивать различные модели регрессии?"), однако это множество ответов может выходить за рамки курса. Знание методов, выходящих за рамки курса, а также способность сравнивать различные методы, приветствуется на экзамене.

#### **10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)**

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

#### **11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)**

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Компьютерный класс.

#### **12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья**

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;

- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 02.04.02 "Фундаментальная информатика и информационные технологии" и магистерской программе "Наука о Данных".



Приложение 2  
к рабочей программе дисциплины (модуля)  
Б1.В.05 Анализ больших данных

**Перечень литературы, необходимой для освоения дисциплины (модуля)**

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Наука о Данных

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

**Основная литература:**

1. Лемешко Б. Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов и др. - М.: НИЦ ИНФРА-М, 2015. - 890 с. ISBN 978-5-16-103267-1. - Режим доступа: <http://znanium.com/bookread2.php?book=515227>
2. Симушкин, С. В. Методы теории вероятностей : учебное пособие / С. В. Симушкин. - Санкт-Петербург : Лань, 2020. - 548 с. - ISBN 978-5-8114-3442-8. - Текст: электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/110911> (дата обращения: 27.07.2020). - Режим доступа: для авториз. пользователей.
3. Статистические методы анализа данных: Учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга [и др.]; под общ. ред. д-ра экон. наук, проф. Л.И. Ниворожкиной. - М.: РИОР: ИНФРА-М, 2016. - 333 с. - (Высшее образование). - [www.dx.doi.org/10.12737/21064](http://www.dx.doi.org/10.12737/21064). - Режим доступа: <http://znanium.com/bookread2.php?book=556760>

**Дополнительная литература:**

1. Кузнецов, Б.Т. Математические методы финансового анализа: учеб. пособие для студентов вузов, обучающихся по специальностям 061800 'Математические методы в экономике', 060400 'Финансы и кредит' / Б.Т. Кузнецов. - М. : ЮНИТИ-ДАНА, 2017. - 159 с. - ISBN 978-5-238-00977-1. - Режим доступа: <http://znanium.com/catalog/product/1028520>
2. Райгородский А.М. Комбинаторика и теория вероятностей: Учебное пособие / А.М. Райгородский - Долгопрудный: Интеллект, 2013. - 104 с. ISBN 978-5-91559-147-8 - Режим доступа: <http://znanium.com/bookread2.php?book=510484>
3. Статистический анализ данных в MS Excel : учеб. пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. ? М. : ИНФРА-М, 2019. ? 320 с. ? (Высшее образование: Бакалавриат). ? [www.dx.doi.org/10.12737/2842](http://www.dx.doi.org/10.12737/2842). - Режим доступа: <http://znanium.com/catalog/product/987337>
4. Методы, модели, средства хранения и обработки данных : учебник / Э.Г. Дадян, Ю.А. Зеленков. - М. : Вузовский учебник : ИНФРА-М, 2017. - 168 с. - Режим доступа: <http://znanium.com/bookread2.php?book=543943>



Приложение 3  
к рабочей программе дисциплины (модуля)  
Б1.В.05 Анализ больших данных

**Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем**

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Наука о Данных

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.