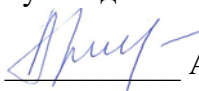
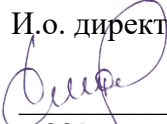




МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
«Дальневосточный федеральный университет»  
(ДФУ)

**ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)**

«СОГЛАСОВАНО»  
Руководитель ОП  
 Артемьева И.Л.

«Утверждаю»  
И.о. директора департамента  
 Смагин С.В.  
«03» марта 2023 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
Основы аналитики больших объёмов данных  
Направление подготовки 09.04.04 Программная инженерия  
(Разработка программно-информационных систем)  
Форма подготовки: очная

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 09.04.04 **Программная инженерия**, утвержденного приказом Министерства образования и науки РФ от 19.09.2017 № 932 (с изменениями и дополнениями).

Рабочая программа обсуждена на заседании департамента программной инженерии и искусственного интеллекта, протокол № 3.0 от «2» марта 2023 г

И.о. директора департамента программной инженерии и искусственного интеллекта Смагин С.В.

Составитель (ли): доцент департамента ПИИИИ к.т.н. Смагин С.В.

Владивосток  
2023

**Оборотная сторона титульного листа РПД**

I. **Рабочая программа пересмотрена на заседании департамента:**  
Протокол от « \_\_\_\_ » \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Директор департамента \_\_\_\_\_  
(подпись) (И.О.Фамилия)

II. **Рабочая программа пересмотрена на заседании департамента:**  
Протокол от « \_\_\_\_ » \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Директор департамента \_\_\_\_\_  
(подпись) (И.О.Фамилия)

## Аннотация дисциплины Основы аналитики больших объёмов данных

Общая трудоемкость дисциплины 3 з.е. (108 час.). Дисциплина является обязательной дисциплиной вариативной части учебного плана: Б1.В.01.03 , изучается на 2 курсе (4 семестр) и завершается экзаменом в 4 семестре. Учебным планом предусмотрено проведение лекционных занятий в объеме 10 часов, практических 12 час. (в том числе интерактивных 12 часов), а также выделены часы на самостоятельную работу студента - 86 час., в том числе 36 час. на подготовку к экзамену.

**Цель дисциплины** – изучение современных методов интеллектуального анализа данных, а также способов формирования и анализа оценок их внешних и внутренних свойств.

### **Задачи дисциплины:**

- Изучение алгоритмов обработки данных, применяемых для случая больших данных
- Изучение особенностей этих алгоритмов и методов их применения.
- Изучение методов сравнения алгоритмов и подготовки альтернативных решений.

Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы, характеризуют формирование следующих компетенций:

Наименование категории (группы) компетенций	Код и наименование компетенции (результат освоения)	Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
Производственно-технологический	ПК-4. Способен создавать программное обеспечение для анализа и обработки информации	ПК-4.1. Демонстрирует знание методов создания программного обеспечения для анализа и обработки информации	<u>Знает</u> внешние и внутренние свойства алгоритмов обработки больших объемов данных, а также основные достижения современных технологий в этой области <u>Умеет</u> сформулировать основные требования к свойствам алгоритма анализа больших данных <u>Владеет</u> схемой компьютерного эксперимента для вычисления оценок свойств алгоритмов анализа больших данных

		ПК-4.2. Использует методы создания программного обеспечения для анализа и обработки информации	<i>Знает</i> критерии выбора наиболее подходящего метода распознавания и обработки данных для произвольной предметной области <i>Умеет</i> провести компьютерный эксперимент для вычисления оценок свойств алгоритмов анализа больших данных <i>Владеет</i> способностью сравнить два алгоритма на основе их внешних и внутренних оценок
		ПК-4.3. Применяет методы создания программного обеспечения для анализа и обработки информации, требуемых в профессиональной деятельности	<i>Знает</i> современные платформы, среды, программные средства и библиотеки, в которых реализованы наиболее популярные и эффективные алгоритмы анализа больших данных <i>Умеет</i> оценить алгоритм анализа больших данных с точки зрения требуемого времени на его построение и качества его работы. <i>Владеет</i> практическими навыками работы в Microsoft Azure и Google Colaboratory
научно-исследовательский	ПК-11. Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем, основанных на знаниях, по обеспечению требуемых критериев эффективности и качества функционирования	ПК-11.1. Выбирает и разрабатывает программные компоненты систем, основанных на знаниях	<i>Знает</i> основные критерии эффективности и качества функционирования системы, основанной на знаниях: точность, релевантность, достоверность, целостность, быстрота решения задач, надежность, защищенность функционирования систем, основанных на знаниях; методы, языки и программные средства разработки программных компонентов систем, основанных на знаниях <i>Умеет</i> выбирать, адаптировать, разрабатывать и интегрировать программные компоненты систем, основанных на знаниях, с учетом основных критериев эффективности и качества функционирования <i>Владеет</i> навыками разработки программных компонентов систем, основанных на знаниях
		ПК-11.2. Проводит экспериментальную проверку работоспособности	<i>Знает</i> методы постановки задач, проведения и анализа тестовых и экспериментальных испытаний работоспособности систем,

		систем, основанных на знаниях	основанных на знаниях <u>Умеет</u> ставить задачи и проводить тестовые и экспериментальные испытания работоспособности систем, основанных на знаниях, анализировать результаты и вносить изменения <u>Владеет</u> навыками проведения и анализа тестовых и экспериментальных испытаний работоспособности систем, основанных на знаниях
производственно-технологический	ПК-13. Способен управлять проектами по созданию, поддержке и использованию систем бизнес-аналитики в организации	ПК-13.1. Разрабатывает архитектуры систем бизнес-аналитики для различных предметных областей	<u>Умеет</u> задачи и роль систем бизнес-аналитики в поддержке принятия решений в процессе управления организацией, принципы построения систем бизнес-аналитики <u>Владеет</u> моделировать и анализировать процессы принятия управленческих решений и разрабатывать требования к системам бизнес-анализа в различных сферах деятельности <u>Владеет</u> навыками разработки требований к системам бизнес-анализа в различных сферах деятельности
		ПК-13.2. Выбирает комплексы методов и инструментальных средств бизнес-аналитики для решения задач в зависимости от особенностей предметной области	<u>Знает</u> методы, технологии, инструменты и платформы бизнес-аналитики; методы анализа данных, используемых в системах бизнес-аналитики для принятия решений <u>Умеет</u> применять методы, инструменты и цифровые платформы анализа данных при проектировании и построении систем бизнес-аналитики <u>Владеет</u> методами, технологиями, инструментами и платформами бизнес-аналитики для решения поставленных задач
		ПК-13.3. Осуществляет руководство проектом по построению системы бизнес-аналитики в организации	<u>Знает</u> методологию и принципы руководства проектом по созданию, поддержке и использованию систем бизнес-аналитики в организации <u>Умеет</u> решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования системы бизнес-аналитики <u>Владеет</u> навыками оценивания результатов внедрения системы бизнес-аналитики в организации и разработки рекомендаций по

			совершенствованию, и развитию системы
организационно-управленческий	ПК-17. Способен осуществлять руководство по созданию и развитию систем и комплексов обработки данных, в том числе больших данных, для корпоративных и государственных заказчиков	ПК-17.1. Осуществляет личное участие в проектах в роли архитектора центра обработки данных, технологического эксперта, специалиста по предпродажам	<u>Знает</u> существующие и перспективные структуры центров обработки данных; действующую нормативную базу в области проектирования и строительства центров обработки данных <u>Умеет</u> формулировать технические задания по формированию аппаратного обеспечения и программных комплексов центра обработки данных. <u>Владеет</u> навыками проведения переговоров с подрядчиками и поставщиками оборудования и программного обеспечения для центра обработки данных
		ПК-17.2. Планирует и осуществляет технологическое развитие центров обработки данных, наращивание и поддержание технологических мощностей и компетенций подразделений	<u>Знает</u> принципы разработки проектной документации центра обработки данных; системную архитектуру серверного оборудования и систем хранения данных, цифровых платформ анализа данных <u>Умеет</u> ставить задачи, планировать и контролировать выполнение работ сотрудниками центра обработки данных <u>Владеет</u> навыками работы с UNIX-подобными системами
		ПК-17.3. Участвует в создании (модернизации) общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области	<u>Знает</u> принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта <u>Умеет</u> применять принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных. <u>Владеет</u> навыками применения принципов и методов хранения

		искусственного интеллекта	наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта
--	--	---------------------------	--

## I. Цели и задачи освоения дисциплины:

**Цель:** изучение теоретических и практических аспектов технологий, в основе которых лежат принципы получения, преобразования, распределенного хранения и обработки, а также анализа больших объемов данных.

### Задачи:

Обзор основных понятий больших данных (Big Data).

2. Изучение основных методов аналитики больших данных.

3. Изучение современных технологий и инструментальных средств по работе с большими данными (Hadoop, MapReduce, Spark, NoSQL, язык R).

Для успешного изучения дисциплины у обучающихся должны быть сформированы предварительные компетенции:

- способность к самоорганизации и самообразованию;
- владение основными концепциями, принципами, теориями и фактами, связанными с информатикой;
- способность осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий;
- способность к формализации в своей предметной области с учетом ограничений используемых методов исследования;
- способность формализовать предметную область программного проекта и разработать спецификации для компонентов программного продукта.

Профессиональные компетенции выпускников и индикаторы их достижения:

Наименование категории (группы) компетенций	Код и наименование компетенции (результат освоения)	Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
Производственно-технологический	ПК-4. Способен создавать программное обеспечение для анализа и обработки информации	ПК-4.1. Демонстрирует знание методов создания программного обеспечения для анализа и обработки информации	<i>Знает</i> внешние и внутренние свойства алгоритмов обработки больших объемов данных, а также основные достижения современных технологий в этой области <i>Умеет</i> сформулировать основные требования к свойствам алгоритма анализа больших данных <i>Владеет</i> схемой компьютерного эксперимента для вычисления оценок свойств алгоритмов анализа больших данных



		ПК-4.2. Использует методы создания программного обеспечения для анализа и обработки информации	<i>Знает</i> критерии выбора наиболее подходящего метода распознавания и обработки данных для произвольной предметной области <i>Умеет</i> провести компьютерный эксперимент для вычисления оценок свойств алгоритмов анализа больших данных <i>Владеет</i> способностью сравнить два алгоритма на основе их внешних и внутренних оценок
		ПК-4.3. Применяет методы создания программного обеспечения для анализа и обработки информации, требуемых в профессиональной деятельности	<i>Знает</i> современные платформы, среды, программные средства и библиотеки, в которых реализованы наиболее популярные и эффективные алгоритмы анализа больших данных <i>Умеет</i> оценить алгоритм анализа больших данных с точки зрения требуемого времени на его построение и качества его работы. <i>Владеет</i> практическими навыками работы в Microsoft Azure и Google Colaboratory
научно-исследовательский	ПК-11. Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем, основанных на знаниях, по обеспечению требуемых критериев эффективности и качества функционирования	ПК-11.1. Выбирает и разрабатывает программные компоненты систем, основанных на знаниях	<i>Знает</i> основные критерии эффективности и качества функционирования системы, основанной на знаниях: точность, релевантность, достоверность, целостность, быстрота решения задач, надежность, защищенность функционирования систем, основанных на знаниях; методы, языки и программные средства разработки программных компонентов систем, основанных на знаниях <i>Умеет</i> выбирать, адаптировать, разрабатывать и интегрировать программные компоненты систем, основанных на знаниях, с учетом основных критериев эффективности и качества функционирования <i>Владеет</i> навыками разработки программных компонентов систем, основанных на знаниях
		ПК-11.2. Проводит экспериментальную проверку работоспособности	<i>Знает</i> методы постановки задач, проведения и анализа тестовых и экспериментальных испытаний работоспособности систем,

		систем, основанных на знаниях	основанных на знаниях <u>Умеет</u> ставить задачи и проводить тестовые и экспериментальные испытания работоспособности систем, основанных на знаниях, анализировать результаты и вносить изменения <u>Владеет</u> навыками проведения и анализа тестовых и экспериментальных испытаний работоспособности систем, основанных на знаниях
производственно-технологический	ПК-13. Способен управлять проектами по созданию, поддержке и использованию систем бизнес-аналитики в организации	ПК-13.1. Разрабатывает архитектуры систем бизнес-аналитики для различных предметных областей	<u>Умеет</u> задачи и роль систем бизнес-аналитики в поддержке принятия решений в процессе управления организацией, принципы построения систем бизнес-аналитики <u>Владеет</u> моделировать и анализировать процессы принятия управленческих решений и разрабатывать требования к системам бизнес-анализа в различных сферах деятельности <u>Владеет</u> навыками разработки требований к системам бизнес-анализа в различных сферах деятельности
		ПК-13.2. Выбирает комплексы методов и инструментальных средств бизнес-аналитики для решения задач в зависимости от особенностей предметной области	<u>Знает</u> методы, технологии, инструменты и платформы бизнес-аналитики; методы анализа данных, используемых в системах бизнес-аналитики для принятия решений <u>Умеет</u> применять методы, инструменты и цифровые платформы анализа данных при проектировании и построении систем бизнес-аналитики <u>Владеет</u> методами, технологиями, инструментами и платформами бизнес-аналитики для решения поставленных задач
		ПК-13.3. Осуществляет руководство проектом по построению системы бизнес-аналитики в организации	<u>Знает</u> методологию и принципы руководства проектом по созданию, поддержке и использованию систем бизнес-аналитики в организации <u>Умеет</u> решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования системы бизнес-аналитики <u>Владеет</u> навыками оценивания результатов внедрения системы бизнес-аналитики в организации и разработки рекомендаций по

			совершенствованию, и развитию системы
организационно-управленческий	ПК-17. Способен осуществлять руководство по созданию и развитию систем и комплексов обработки данных, в том числе больших данных, для корпоративных и государственных заказчиков	ПК-17.1. Осуществляет личное участие в проектах в роли архитектора центра обработки данных, технологического эксперта, специалиста по предпродажам	<u>Знает</u> существующие и перспективные структуры центров обработки данных; действующую нормативную базу в области проектирования и строительства центров обработки данных <u>Умеет</u> формулировать технические задания по формированию аппаратного обеспечения и программных комплексов центра обработки данных. <u>Владеет</u> навыками проведения переговоров с подрядчиками и поставщиками оборудования и программного обеспечения для центра обработки данных
		ПК-17.2. Планирует и осуществляет технологическое развитие центров обработки данных, наращивание и поддержание технологических мощностей и компетенций подразделений	<u>Знает</u> принципы разработки проектной документации центра обработки данных; системную архитектуру серверного оборудования и систем хранения данных, цифровых платформ анализа данных <u>Умеет</u> ставить задачи, планировать и контролировать выполнение работ сотрудниками центра обработки данных <u>Владеет</u> навыками работы с UNIX-подобными системами
		ПК-17.3. Участвует в создании (модернизации) общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области	<u>Знает</u> принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта <u>Умеет</u> применять принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных. <u>Владеет</u> навыками применения принципов и методов хранения

		искусственного интеллекта	наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта
--	--	---------------------------	--

## II. Трудоемкость дисциплины и виды учебных занятий по дисциплине

Общая трудоемкость дисциплины составляет 3 зачётных единиц (108 академических часа).

Структура дисциплины:

Форма обучения – очная

№	Наименование раздела дисциплины	Семестр	Количество часов по видам учебных занятий и работы обучающегося						Формы промежуточной аттестации
			Лек	Лаб	Пр	ОК	СР	Контроль	
1	Введение в большие данные	4	1		2				УО-1, УО-4, ПР-6 экзамен
2	Жизненный цикл аналитики данных	4	2		2				
3	Высокопроизводительные вычисления	4	2		2				
4	Масштабирование и многоуровневое хранение данных	4	2		3		48	36	
5	Визуализация данных и результатов анализа	4	2		3		3		
6	Сложные методы аналитики	4	2						
7	Анализ текста	4	1						
	Итого:		10		12		50	36	экзамен

\*онлайн курс

### **III. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА**

#### **Лекции (10 час)**

##### **Тема 1. Введение в большие данные (1 час.)**

Рассматриваются понятия: большие данные, причины появления больших данных, примеры возможностей для бизнеса, принцип трех «V» (volume, velocity, variety), рост объемов информации, основные типы данных (структурированные, полуструктурированные, квазиструктурированные и неструктурированные).

##### **Тема 2. Жизненный цикл аналитики данных (1 час.)**

Рассматриваются понятия: Business Intelligence и Big Data, жизненный цикл аналитики данных, Data Science, хранилище данных, интеллектуальный анализ данных, предиктивное моделирование, BI-платформа, роли для создания проекта по аналитике данных, инструменты получения и обмена данными.

##### **Тема 3. Высокопроизводительные вычисления (1 час.)**

Рассматриваются понятия: распределенные вычисления на нескольких серверах, вычислительная парадигма MapReduce, проект Apache Hadoop и его экосистема, Apache Spark и его компоненты, вычисления в реальном времени, Apache Storm, Flink, HDFS (Hadoop Distributed File System).

##### **Тема 4. Масштабирование и многоуровневое хранение данных (2 час.)**

Рассматриваются понятия: CAP-теорема, одноранговая сеть, парадигма NoSQL, классификация NoSQL баз данных, масштабирование (вертикальное, горизонтальное), репликация, шардинг (вертикальный, горизонтальный).

##### **Тема 5. Визуализация данных и результатов анализа (2 час.)**

Рассматриваются понятия: типы, задачи и техники визуализации данных, язык R (отличия, основные возможности).

##### **Тема 6. Сложные методы аналитики (2 час.)**

Рассматриваются понятия: классификация задач анализа (Text Mining, Data Mining, Web Mining, Social Mining), статистические методы анализа данных, применение машинного обучения в аналитике.

##### **Тема 7. Анализ текста (1 час.)**

Рассматриваются понятия: поисковые механизмы (Lucene, Solr, Elasticsearch), алгоритм Work2Vec.

### **IV. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА**

## **ПРАКТИЧЕСКИЕ ЗАНЯТИЯ (12 час.)**

### **Практическое занятие 1. Выбор предметной области (2 час.)**

Поставленная задача не привязана к какой-либо конкретной предметной области. Предполагается отойти от принципа выполнения заранее поставленных и четко сформулированных задач, чтобы предоставить исполнителю гибкость и возможность творческого подхода выполнения. Таким образом, исполнителю предоставляется возможность самостоятельного выбора интересующей его прикладной области, над которой в рамках курса будет проводиться работа.

### **Практическое занятие 2. Формирование набора данных (2 час.)**

Во время выполнения задания может потребоваться работать с информацией разного типа. Традиционно принято выделять четыре типа данных: структурированные, полуструктурированные, квазиструктурированные и не структурированные данные. Исполнитель самостоятельно выбирает тип данных, с которым в дальнейшем будет работать, но требуется принимать во внимание, что поскольку в курсе рассматриваются подходы и технологии обработки именно большого объема данных, то для выбранной прикладной области рекомендуется иметь для проведения анализа не менее 2 Гб структурированных или полуструктурированных данных, если не используются методы анализа неструктурированного контента. В случае, если используются методы анализа неструктурированного контента, такого как изображения, аудио- и видеозаписи, то рекомендуемый минимальный объем информации – 5 Гб.

### **Практическое занятие 3. Архитектура проектируемой системы (2 час.)**

Задание подразумевает создание программного решения, позволяющего автоматически или полуавтоматически решать сформулированные задачи анализа. В основе решения может быть заложена относительно простая, но функциональная и расширяемая модульная схема. Стоит отметить, что программное решение должно обладать хорошей производительностью, гибким масштабированием, быть распределенным и гарантировать надежность передачи данных между узлами системы. Еще одной важной особенностью рассматриваемой архитектуры является возможность гибкой настройки проводимых в системе аналитик.

### **Практическое занятие 4. Хранение и обработка данных (3 час.)**

Хранение сформированного набора данных или набора данных, который прошел предварительную очистку и готов поступить на обработку, предполагается осуществлять в базе данных. Модуль обработки данных

является одним из центральных модулей программной системы анализа. В нем заложена основная логика получения ответа на поставленную задачу анализа. Выбор используемых технологий и методов решения зависит именно от задачи. Для одного типа задач хорошо подходит использование методов машинного обучения и нейронных сетей, а для другого типа задач идеальным образом становится простое решение с помощью SQL-подобных запросов.

### **Практическое занятие 5. Визуализация результатов (3 час.)**

Результаты проведенных исследований над данными должны быть представлены в виде набора графиков и диаграмм, наглядно изображающих полученные выводы. Возможно использование подходов изображения результатов в виде инфографики или облака тегов. Для визуализации результатов аналитики разумно применение языка R. Он может быть использован для обработки «сырых» результатов анализа, объем которых не превышает 300 Мб, так как программы, созданные на языке R, являются относительно медленными и не масштабируемыми, хотя на сегодняшний день и есть поддержка языка R в Spark для исполнения на кластере. Программы на языке R могут быть использованы для быстрого и наглядного изображения промежуточных результатов, на основе которых делается выбор дальнейшего направления движения при ответе на поставленный в задаче вопрос.

### **Лабораторные работы (0 час.)**

## **V. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА**

№ п/п	Контролируемые разделы/темы дисциплины	Код индикатора достижения компетенции	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
1.	Темы 1-9	ПК-4.1. Демонстрирует знание методов создания программного обеспечения для анализа и обработки информации.	Знает способы получения оценок моделей и методов машинного обучения, а также критериев их сравнения. Умеет поставить задачу машинного обучения в заданной	УО-1, УО-4	Экзамен Вопросы 1-30

			<p>предметной области и предложить метод ее решения.</p> <p>Владеет навыками работы с платформами, средами и программными средствами, в библиотеках которых реализованы методы машинного обучения.</p>		
		ПК-4.2. Использует методы создания программного обеспечения для анализа и обработки информации.	<p>Знает способы выбора наиболее подходящего метода машинного обучения для заданной предметной области на основе оценки и анализа свойств существующих методов.</p> <p>Умеет проанализировать заданную предметную область на предмет возможности и необходимости применения в ней методов машинного обучения.</p>		
			Владеет навыками применения заданного метода машинного обучения (на основе существующих библиотек), а также получения результата (на основе заданного набора данных).	ПР-6	
		ПК-4.3. Применяет методы создания программного обеспечения для анализа и обработки информации,	Знает основные репозитории, в которых хранятся наборы данных для машинного обучения, и способы их использования.		



		требуемых в профессиональной деятельности.	<p>Умеет обработать заданный набор данных различными методами машинного обучения и визуализировать результат.</p> <p>Владеет способностью интерпретации полученных результатов и их улучшения посредством настройки значений параметров.</p>		
2.	Темы 10-15	<p>ПК-13.1. Разрабатывает архитектуру систем бизнес-аналитики для различных предметных областей</p> <p>ПК-13.2. Выбирает комплексы методов и инструментальных средств бизнес-аналитики для решения задач</p>	<p>Знает задачи и роль систем бизнес-аналитики в поддержке принятия решений в процессе управления организацией, принципы построения систем бизнес-аналитики</p> <p>Умеет моделировать и анализировать процессы принятия управленческих решений и разрабатывать требования к системам бизнес-анализа в различных сферах деятельности</p> <p>Владеет навыками разработки требований к системам бизнес-анализа в различных сферах деятельности</p> <p>Знает методы, технологии, инструменты и платформы бизнес-аналитики; методы анализа данных, используемых в системах бизнес-</p>	УО-1, УО-4	Экзамен Вопросы 31-59

		<p>в зависимости от особенностей предметной области</p>	<p>аналитики для принятия решений</p> <p>Умеет применять методы, инструменты и цифровые платформы анализа данных при проектировании и построении систем бизнес-аналитики</p> <p>Владеет методами, технологиями, инструментами и платформами бизнес-аналитики для решения поставленных задач</p>	<p>ПР-6</p>	
		<p>ПК-13.3. Осуществляет руководство проектом по построению системы бизнес-аналитики в организации</p>	<p>Знает методологию и принципы руководства проектом по созданию, поддержке и использованию систем бизнес-аналитики в организации</p> <p>Умеет решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования системы бизнес-аналитики</p> <p>Владеет навыками оценивания результатов внедрения системы бизнес-аналитики в организации и разработки рекомендаций по совершенствованию, и развитию системы</p>		

## VI. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Самостоятельная работа определяется как индивидуальная или коллективная учебная деятельность, осуществляемая без непосредственного руководства педагога, но по его заданиям и под его контролем. Самостоятельная работа – это познавательная учебная деятельность, когда последовательность мышления студента, его умственных и практических операций и действий зависит и определяется самим студентом.

Самостоятельная работа студентов способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня, что в итоге приводит к развитию навыка самостоятельного планирования и реализации деятельности.

Целью самостоятельной работы студентов является овладение необходимыми компетенциями по своей специальности, опытом творческой и исследовательской деятельности.

Формы самостоятельной работы студентов:

- работа с основной и дополнительной литературой, Интернет ресурсами;
- самостоятельное ознакомление с лекционным материалом, представленным на электронных носителях, в библиотеке образовательного учреждения;
- решение задач;
- подготовка к экзамену;
- другие виды деятельности, организуемые и осуществляемые образовательным учреждением и органами студенческого самоуправления.

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Основы аналитики больших объёмов данных» включает в себя план-график выполнения самостоятельной работы по дисциплине.

## План-график выполнения самостоятельной работы по дисциплине

№ п/п	Дата/сроки выполнения	Вид самостоятельной работы	Примерные нормы времени на выполнение	Форма контроля
1	7 неделя обучения	Выбор предметной области	4 часа	Собеседование
2	8-10 неделя обучения	Формирование набора данных	10 часов	Собеседование
3	10-13 неделя обучения	Архитектура проектируемой системы	10 часов	Проверка отчетов, собеседование
4	14-15 неделя обучения	Хранение и обработка данных	12 часов	Проверка отчетов, собеседование
5	16-17 неделя обучения	Визуализация результатов	14 часов	Проверка отчетов, собеседование
Итого:			50 часов	

Самостоятельная работа по дисциплине включает в себя подготовку к практическим занятиям (изучение литературы), выполнению практических заданий, и подготовку к промежуточной аттестации по дисциплине.

Рекомендуется использовать различные возможности работы с литературой: фонды научной библиотеки ДВФУ (<http://www.dvfu.ru/library/>) и других ведущих вузов страны, а также доступных для использования научно-библиотечных систем.

## VII. СПИСОК ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Основная литература

*(электронные и печатные издания)*

1. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. – М.: Альпина Пабли., 2016. – 461 с.: 84x108 1/16 ISBN 978-5-9614-5032-3 <http://znanium.com/catalog/product/551044>  
<http://www.studentlibrary.ru/book/ISBN9785961450323.html>
1. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. Language Arts & Disciplines, 2013. – 599 с.

2. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб: Питер, 2017. – 336 стр.
3. Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия, алгоритмы Data Mining с использованием R. – 2017. – Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>

**Дополнительная литература**  
(печатные и электронные издания)

4. Уайт Т. Nadoor: Подробное руководство. СПб: Питер, 2013. – 672 с.
5. Лэм Ч. Nadoor в действии. Москва: ДМК Пресс, 2012. – 426 с.
6. Фаулер М., Прамодкумар Дж. Садаладж. NoSQL: новая методология разработки нереляционных баз данных. – М.: «Вильямс», 2013. – 192 с.
7. Смородин В.В., Волкова Е.В., Алиев А. А. От хранения данных к управлению информацией. – СПб.: Изд-во Питер, 2010. – 528 с.
8. Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми образами. – Wiley Publishing, Inc. 2013.
9. Храмов Д.А. Сбор данных в Интернете на языке R, 2016. – 282 с.
- 10.
11. Nadoor в действии / Чак Лэм – М. : ДМК Пресс, 2012. – 424 с. – ISBN 978-5-94074-785-7  
<http://www.studentlibrary.ru/book/ISBN9785940747857.html>
12. Вагин В.Н. Достоверный и правдоподобный вывод в интеллектуальных системах: учебное пособие. – М.: Физматлит, 2008. – 704 с.  
<http://znanium.com/go.php?id=544735>
13. Низаметдинов Ш.У., Румянцев В.П. Анализ данных: учебное пособие. – М.: НИЯУ «МИФИ», 2012. – 288 с. ISBN 978-5-7262-1687-4.  
<http://znanium.com/bookread2.php?book=567083>
14. Болотова Л.С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях: учебник / ФГБОУ ВПО РГУИТП; ФГАУ ГНИИ ИТТ «Информика». – М.: Финансы и статистика, 2012. – 664 с.: ил.  
<http://www.studentlibrary.ru/book/ISBN9785279035304.html>

**Перечень ресурсов информационно-телекоммуникационной сети  
«Интернет»**

1. <http://machinelearning.ru/> MachineLearning.ru Профессиональный информационно-аналитический ресурс, посвященный машинному

обучению, распознаванию образов и интеллектуальному анализу данных.

2. <http://www.inftech.webservis.ru/it/database/datamining/ar2.html> Дюк В.А. Data Mining – интеллектуальный анализ данных.
3. <http://kek.ksu.ru/EOS/dm.pdf> Степанов Р.Г. Технология Data Mining: Интеллектуальный анализ данных / Казань, 2008.

### **Перечень информационных технологий и программного обеспечения**

Для выполнения лабораторных работ требуется следующее программное обеспечение: Microsoft Excel, RGui, RStudio.

## **VIII. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

### **Перечень программного обеспечения:**

#### **Лицензионное программное обеспечение:**

AutoCAD;  
Autodesk 3DS Max;  
Microsoft Visio;  
SPSS Statistics Premium Campus Edition;  
MathCad Education University Edition;  
Microsoft Office 365;  
Office Professional Plus 2019;  
Photoshop CC for teams All Apps AL;  
SolidWorks Campus 500;  
Windows Edu Per Device 10 Education;  
КОМПАС 3D;  
Microsoft Teams

#### **Свободно распространяемое программное обеспечение:**

Adobe Reader DC 2015.020 - пакет программ для просмотра электронных публикаций в формате PDF:

[http://www.images.adobe.com/content/dam/acom/en/legal/licenses-terms/pdf/PlatformClients\\_PC\\_WWEULA-en\\_US-20150407\\_1357.pdf](http://www.images.adobe.com/content/dam/acom/en/legal/licenses-terms/pdf/PlatformClients_PC_WWEULA-en_US-20150407_1357.pdf);

ArgoUML - программный инструмент моделирования UML:

<http://argouml.tigris.org>;

Dia - пакет программ для создания диаграмм в виде блок-схем алгоритмов программ, древовидных схем, статических структур UML, баз данных, диаграмм сущность-связь и др. диаграмм:

[https://portableapps.com/support/portable\\_app#using](https://portableapps.com/support/portable_app#using);

DiagramDesigner - пакет программ для создания потоковых диаграмм, диаграмм классов UML, иллюстраций и др. диаграмм:

<https://www.fosshub.com/Diagram-Designer.html#clickToStartDownload>;

IrfanView - пакет программ для просмотра (воспроизведения) графических, видео- и аудиофайлов <http://www.irfanview.com/eula.htm>;

LibreOffice - офисный пакет:

<http://www.libreoffice.org/about-us/licenses/>;

Maxima – система для работы с символьными и численными выражениями <http://maxima.sourceforge.net/maximalist.html>;

Project Libre - аналог программной системы управления проектами Microsoft Project для стационарного компьютера:

<https://континентсвободы.рф:/офис/проекты/projectlibre-система-управления-проектами.html>;

Python - система программирования - динамический интерактивный объектно-ориентированный язык программирования:

<https://python.ru.uptodown.com/windows/download>;

Ramus Educational - пакет программ для разработки и моделирования бизнес-процессов в виде диаграмм IDEF0 и DFD:

<https://www.obnovisoft.ru/ramus-educational>;

Scilab –система - язык программирования высокого уровня, рассчитанный на научные расчеты: <http://www.scilab.org/scilab/license>;

WhiteStarUML –программный инструмент моделирования UML, полученный из StarUML, совместимый с Windows 7-10:

<https://github.com/StevenTCramer/WhiteStarUml/blob/master/staruml/deploy/License.txt/>;

WinDjView – программа для просмотра электронных публикаций в формате DJV и DjVu: <https://windjview.sourceforge.io/ru/>.

Успешное освоение дисциплины предполагает активную работу студентов на всех занятиях аудиторной формы: лекциях и лабораторных

работах, выполнение аттестационных мероприятий. В процессе изучения дисциплины студенту необходимо ориентироваться на проработку лекционного материала, подготовку к лабораторным работам.

Промежуточной аттестацией по дисциплине «Основы аналитики больших объёмов данных» является экзамен в 4 семестре.

Студент считается аттестованным по дисциплине при условии выполнения всех лабораторных работ и сдачи экзамена.

## **IX. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

Сведения о материально-техническом обеспечении ОПОП, включая информацию о наличии оборудованных учебных кабинетов, объектов для проведения практических занятий и самостоятельной работы обучающихся с перечнем основного оборудования, объектов физической культуры и спорта, программного обеспечения представлены в виде таблицы в Справке об МТО.