



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

Ралин А.Ю.
(ФИО)

«УТВЕРЖДАЮ»

Директор департамента

Пустовалов Е.В.
(ФИО)
«01» марта 2022



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Большие данные

Направление подготовки 09.03.02 Информационные системы и технологии
(Информационные системы и технологии)

Форма подготовки очная

курс 4 семестр 7

лекции 16 час.

практические занятия 0 час.

лабораторные работы 34 час.

в том числе с использованием МАО лек. 0 / пр. 0 / лаб. 18 час.

всего часов аудиторной нагрузки 50 час.

в том числе с использованием МАО 18 час.

самостоятельная работа 94 час.

в том числе на подготовку к экзамену 0 час.

контрольные работы (количество) не предусмотрены

курсовая работа / курсовой проект не предусмотрены

зачет 7 семестр

экзамен не предусмотрен

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 09.03.02 Информационные системы и технологии, утвержденного приказом Министерства образования и науки РФ от 19.07.2017 № 926 (с изменениями и дополнениями).

Рабочая программа обсуждена на заседании департамента информационных и компьютерных систем, протокол № 7 от 25 февраля 2022 г.

Директор департамента информационных и компьютерных систем Пустовалов Е.В.

Составитель: д.ф.-м.н., профессор Плотников В.С.

Владивосток
2022

Оборотная сторона титульного листа РПУД

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)

III. Рабочая программа пересмотрена на заседании департамента:

Протокол от « ____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

IV. Рабочая программа пересмотрена на заседании департамента:

Протокол от « ____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

Аннотация к рабочей программе дисциплины «Большие данные»

Дисциплина «Большие данные» предназначена для студентов, обучающихся по направлению подготовки 09.03.02 Информационные системы и технологии, профиль «Информационные системы и технологии», входит в часть, формируемую участниками образовательных отношений Блока 1 «Дисциплины (модули)» учебного плана (индекс Б1.В.ДВ.05.02) и является дисциплиной по выбору.

Общая трудоемкость освоения дисциплины составляет 4 зачетные единицы, 144 часа. Учебным планом предусмотрены лекционные занятия (16 часов), лабораторные работы (34 часа), самостоятельная работа студентов (94 часа). Дисциплина реализуется на 4 курсе в 7 семестре. Форма промежуточной аттестации – зачет.

Цель изучения дисциплины – получение теоретических знаний и практических навыков, необходимых при работе с большими данными (Big Data), сборе и анализе огромных объемов структурированной или неструктурированной информации, разработке моделей данных и получении новых знаний.

Задачи:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- получение навыков применения статистических и математических методов для анализа больших объемов информации;
- освоение современных представлений о принципах построения систем обработки больших данных и информационных систем на их основе, их структуре и основах аппаратного и программного обеспечения.

В результате изучения данной дисциплины у обучающихся формируются следующие **компетенции**.

Универсальные компетенции и индикаторы их достижения:

Наименование категории (группы) универсальных компетенций	Код и наименование универсальной компетенции выпускника	Код и наименование индикатора достижения универсальной компетенции
Разработка и реализация проектов	УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1. Знать: виды ресурсов и ограничений для решения профессиональных задач; основные методы оценки разных способов решения задач; действующее законодательство и правовые нормы, регулирующие профессиональную деятельность. УК-2.2. Уметь: проводить анализ поставленной цели и формулировать задачи,

		<p>которые необходимо решить для ее достижения; анализировать альтернативные варианты для достижения намеченных результатов; использовать нормативно-правовую документацию в сфере профессиональной деятельности.</p> <p>УК-2.3. Владеть: методиками разработки цели и задач проекта; методами оценки потребности в ресурсах, продолжительности и стоимости проекта; навыками работы с нормативно-правовой документацией.</p>
--	--	---

Профессиональные компетенции и индикаторы их достижения:

Задача профессиональной деятельности	Объект или область знания	Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции	Основание (ПС, анализ иных требований, предъявляемых к выпускникам)
Тип задач профессиональной деятельности: научно-исследовательский				
Исследование моделей и методов информационных систем и технологий	информационные системы и технологии	ПК-1. Способность проводить исследования на всех этапах жизненного цикла информационных систем	<p>ПК-1.1. – знает методологию проведения исследований в области информационных систем и технологий</p> <p>ПК-1.2. – умеет выбирать и применять методы исследования на всех этапах жизненного цикла информационных систем</p> <p>ПК-1.3. – владеет навыками проведения исследований на всех этапах жизненного цикла информационных систем</p>	06 Связь, информационные и коммуникационные технологии (в сфере исследования, разработки, внедрения и сопровождения информационных технологий и систем)
Тип задач профессиональной деятельности: производственно-технологический				
Выполнение работ по созданию (модификации) и сопровождению информационных систем	информационные системы и технологии	ПК-4. Способность выполнять работы по созданию (модификации) и сопровождению информационных систем	<p>ПК-4.1. – знает архитектуру, устройство и функционирование современных информационных систем</p> <p>ПК-4.2. – умеет выполнять работы по созданию (модификации) и сопровождению информационных систем</p> <p>ПК-4.3. – владеет навыками создания, модификации и сопровождения информационных систем</p>	06.015 Специалист по информационным технологиям

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Тема 1. Введение в большие данные (2 час.)

Определение больших данных и причины их появления. Примеры возможностей для бизнеса.

Тема 2. Жизненный цикл аналитики данных (2 час.)

Различие между Business Intelligence и Big Data. Понятие жизненного цикла аналитики данных. Роли, необходимые для успешного создания проекта по аналитике данных. Инструменты получения и обмена данными.

Тема 3. Высокопроизводительные вычисления (2 час.)

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.

Тема 4. Масштабирование и многоуровневое хранение данных (3 час.)

Теорема CAP. Парадигма NoSQL. Классификация NoSQL баз данных.

Тема 5. Визуализация данных и результатов анализа (3 час.)

Техники визуализации данных. Язык R.

Тема 6. Сложные методы аналитики (2 час.)

Классификация задач анализа: Text, Data, Web, Social Mining. Статистические методы анализа данных. Применение машинного обучения в аналитике.

Тема 7. Анализ текста (2 час.)

Поисковые механизмы: Lucene, Solr, Elasticsearch. Алгоритм Work2Vec.

II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА И САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Лабораторные работы (34 час.)

Лабораторная работа № 1. Выбор предметной области (6 час.)

Поставленная перед слушателями 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.

Лабораторная работа № 2. Формирование набора данных (8 час.)

Во время выполнения проекта может потребоваться работать с информацией разного типа. Традиционно принято выделять четыре типа данных: структурированные данные, полуструктурированные данные, квазиструктурированные данные 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 – 5Гб.

Лабораторная работа № 3. Архитектура проектируемой системы (8 час.)

Разрабатываемый проект 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.

Лабораторная работа № 4. Хранение и обработка данных (6 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. Таким образом, выбор того, какую технологию использовать, ложится на слушателя.

Лабораторная работа № 5. Визуализация результатов (6 час.)

Результаты проведенных исследований над данными обычно представляются 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 200...300 Мб, так как программы, созданные на нем (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.

Самостоятельная работа (94 час.)

№ п/п	Дата/сроки выполнения	Вид самостоятельной работы	Примерные нормы времени на выполнение	Форма контроля
1	1 неделя	ЛР Выбор предметной области	10 час.	Собеседование
2	2-5 недели	ЛР Формирование набора данных	16 час.	Собеседование
3	6-10 недели	ЛР Архитектура проектируемой системы	20 час.	Проверка отчетов, собеседование
4	11-14 недели	ЛР Хранение и обработка данных	20 час.	Проверка отчетов, собеседование
5	15-17 недели	ЛР Визуализация результатов	16 час.	Зачет
6	18 неделя	Тестирование	12 час.	Тест
Итого			94 час.	

III. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Рекомендации по самостоятельной работе студентов

Самостоятельная работа обучающихся подразумевает обязательную подготовку к лабораторным занятиям (оформление отчетов), изучение основной и дополнительно литературы по дисциплине, подготовку к текущему контролю и промежуточной аттестации в конце семестра, консультации преподавателя.

Рекомендации по работе с литературой

Для более эффективного освоения и усвоения материала рекомендуется ознакомиться с теоретическим материалом по той или иной теме до проведения лабораторного занятия. Всю учебную литературу желательно изучать «под конспект». Цель написания конспекта по дисциплине –

сформировать навыки по поиску, отбору, анализу и формулированию учебного материала. При работе над конспектом обязательно выявляются и отмечаются трудные для самостоятельного изучения вопросы, с которыми уместно обратиться к преподавателю при посещении консультаций, либо в индивидуальном порядке.

Подготовка к лабораторным работам

Подготовку к лабораторной работе студент должен начать с изучения теоретического материала и ознакомления с планом, который отражает содержание предложенной темы. Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы по теме задания, и правильном его выполнении.

В процессе выполнения лабораторной работы студент должен создать требуемый документ с помощью предлагаемого программного средства и выполнить требуемые в задании операции. Задание содержит методические указания по подготовке документа, который должен быть получен в результате выполнения работы. При подготовке следует их внимательно прочесть.

IV. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций	Оценочные средства		
			текущий контроль	промежуточная аттестация	
1	Темы 1-2	УК-2 ПК-1 ПК-4	знает	собеседование УО-1, круглый стол УО-4	Зачет Вопросы 1-20
			умеет	лабораторная работа (ПР-6)	
			владеет	лабораторная работа (ПР-6)	
2	Темы 3-4	УК-2 ПК-1 ПК-4	знает	собеседование УО-1, круглый стол УО-4	Зачет Вопросы 21-45
			умеет	лабораторная работа (ПР-6)	
			владеет	лабораторная работа (ПР-6)	
3	Темы 5-7	УК-2 ПК-1 ПК-4	знает	собеседование УО-1, круглый стол УО-4	Зачет Вопросы 46-59

			умеет	лабораторная работа (ПР-6)	
			владеет	лабораторная работа (ПР-6)	

V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(электронные и печатные издания)

1. 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>
2. Вагин В.Н. Достоверный и правдоподобный вывод в интеллектуальных системах: учебное пособие. – М.: Физматлит, 2008. – 704 с. <http://znanium.com/go.php?id=544735>
3. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. – М.: Альпина Пабл., 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>
4. 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>
5. Hadoop в действии / Чак Лэм – М.: ДМК Пресс, 2012. – 424 с. – ISBN 978-5-94074-785-7 <http://www.studentlibrary.ru/book/ISBN9785940747857.html>

Дополнительная литература

(электронные и печатные издания)

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. Уайт Т. Hadoop: Подробное руководство. СПб: Питер, 2013. – 672 с.
4. Лэм Ч. Hadoop в действии. Москва: ДМК Пресс, 2012. – 426 с.
5. Фаулер М., Прамодкумар Дж. Садаладж. NoSQL: новая методология разработки нереляционных баз данных. – М.: «Вильямс», 2013. – 192 с.

6. Смородин В.В., Волкова Е.В., Алиев А. А. От хранения данных к управлению информацией. – СПб.: Изд-во Питер, 2010. – 528 с.
7. Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми образами. – Wiley Publishing, Inc. 2013.
8. Храмов Д.А. Сбор данных в Интернете на языке R, 2016. – 282 с.
9. Шитиков В.К., Мастицкий С.Э. Классификация, регрессия, алгоритмы Data Mining с использованием R. – 2017. – Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>

Перечень ресурсов информационно-телекоммуникационной сети

Интернет

1. <http://www.inftech.webservis.ru/it/database/datamining/ar2.html> Дюк В.А. Data Mining – интеллектуальный анализ данных.
2. <http://kek.ksu.ru/EOS/dm.pdf> Степанов Р.Г. Технология Data Mining: Интеллектуальный анализ данных / Казань, 2008.
3. <http://machinelearning.ru/> MachineLearning.ru Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.

VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Для изучения учебной дисциплины необходимо вспомнить и систематизировать знания, полученные ранее по данной отрасли научного знания. При изучении материала по учебнику нужно, прежде всего, уяснить существо каждого излагаемого там вопроса. Главное - это понять изложенное в учебнике, а не «заучить». Сначала следует прочитать весь материал темы (параграфа), особенно не задерживаясь на том, что показалось не совсем понятным: часто это становится понятным из последующего. Затем надо вернуться к местам, вызвавшим затруднения и внимательно разобраться в том, что было неясно. Особое внимание при повторном чтении необходимо обратить на формулировки соответствующих определений, формулы и т.п.; в точных формулировках, как правило, существенно каждое слово и очень полезно понять, почему данное положение сформулировано именно так. Однако не следует стараться заучивать формулировки; важно понять их смысл и уметь изложить результат своими словами. Закончив изучение раздела, полезно составить краткий конспект, по возможности не заглядывая в учебник (учебное пособие).

При изучении учебной дисциплины особое внимание следует уделить приобретению навыков решения профессионально-ориентированных задач.

Для этого, изучив материал данной темы, надо сначала обязательно разобраться в решениях соответствующих задач, которые рассматривались на практических занятиях, приведены в учебно-методических материалах, пособиях, учебниках, ресурсах Интернета, обратив особое внимание на методические указания по их решению. Затем необходимо самостоятельно решить несколько аналогичных задач из сборников задач, и после этого решать соответствующие задачи из сборников тестовых заданий и контрольных работ. Закончив изучение раздела, нужно проверить умение ответить на все вопросы программы курса по этой теме (осуществить самопроверку).

Все вопросы, которые должны быть изучены и усвоены, в программе перечислены достаточно подробно. Однако очень полезно составить перечень таких вопросов самостоятельно (в отдельной тетради) следующим образом:

- начав изучение очередной темы программы, выписать сначала в тетради последовательно все перечисленные в программе вопросы этой темы, оставив справа широкую колонку;

- по мере изучения материала раздела (чтения учебника, учебно-методических пособий, конспекта лекций) следует в правой колонке указать страницу учебного издания (конспекта лекции), на которой излагается соответствующий вопрос, а также номер формулы, которые выражают ответ на данный вопрос. В результате в этой тетради будет полный перечень вопросов для самопроверки, который можно использовать и при подготовке к экзамену. Кроме того, ответив на вопрос или написав соответствующую формулу (уравнение), можете по учебнику (конспекту лекций) быстро проверить, правильно ли это сделано, если в правильности своего ответа есть сомнения. Наконец, по тетради с такими вопросами можно установить, весь ли материал, предусмотренный программой, изучен.

Следует иметь в виду, что в различных учебных изданиях материал может излагаться в разной последовательности. Поэтому ответ на какой-нибудь вопрос программы может оказаться в другой главе, но на изучении курса в целом это, конечно, никак не скажется. Указания по выполнению тестовых заданий и контрольных работ приводятся в учебно-методической литературе, в которых к каждой задаче даются конкретные методические указания по ее решению и приводится пример решения.

Для углубленного изучения теоретического материала курса дисциплины рекомендуются использовать основную и дополнительную литературу, указанную в приведенном выше перечне.

Рекомендованные источники доступны обучаемым в научной библиотеке (НБ) ДВФУ (в перечне приведены соответствующие гиперссылки

этих источников), а также в электронной библиотечной системе (ЭБС) IPRbooks (приведены аналогичные гиперссылки).

Для подготовки к зачету определен перечень вопросов, представленный ниже, в материалах фонда оценочных средств дисциплины.

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для проведения учебных занятий по дисциплине, а также для организации самостоятельной работы студентам доступно следующее лабораторное оборудование и специализированные кабинеты, соответствующие действующим санитарным и противопожарным нормам, а также требованиям техники безопасности при проведении учебных и научно-производственных работ:

Наименование оборудованных помещений и помещений для самостоятельной работы	Перечень основного оборудования
г. Владивосток, о. Русский, п. Аякс д.10, корпус L, ауд. L558 учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Парты и стулья Мультимедийное оборудование: проектор BENQ CH100, ноутбук ACER ASPIRE TimeLine 3495
г. Владивосток, о. Русский, п. Аякс д.10, корпус L, ауд. L450 15 мест специализированная лаборатория кафедры КС: Лаборатория администрирования информационных систем	11 компьютеров (системный блок модель - 30AGCT01WW P3+монитором AOC 28" LI2868POU) Специализированная мебель (столы и стулья)
Читальные залы Научной библиотеки ДВФУ с открытым доступом к фонду (корпус А - уровень 10)	Моноблок HP ProOne 400 All-in-One 19,5 (1600x900), Core i3-4150T, 4GB DDR3-1600 (1x4GB), 1TB HDD 7200 SATA, DVD+/-RW, GigEth, Wi-Fi, BT, usb kbd/mse, Win7Pro (64-bit)+Win8.1Pro(64-bit), 1-1-1 Wty Скорость доступа в Интернет 500 Мбит/сек. Рабочие места для людей с ограниченными возможностями здоровья оснащены

Аудитория для самостоятельной работы	дисплеями и принтерами Брайля; оборудованы: портативными устройствами для чтения плоскочечатных текстов, сканирующими и читающими машинами видеоувелечителем с возможностью регуляции цветовых спектров; увеличивающими электронными лупами и ультразвуковыми маркировщиками
--------------------------------------	--

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы пандусами, лифтами, подъемниками, специализированными местами, оснащенными туалетными комнатами, табличками информационно-навигационной поддержки.

VIII. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Текущая аттестация студентов. Текущая аттестация студентов по дисциплине «Большие данные» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация по дисциплине «Большие данные» проводится в форме контрольных мероприятий (защита проекта, тестирование) по оцениванию фактических результатов обучения студентов осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- учебная дисциплина (активность на занятиях, своевременность выполнения различных видов заданий, посещаемость всех видов занятий по аттестуемой дисциплине);
- степень усвоения теоретических знаний;
- уровень овладения практическими умениями и навыками по всем видам учебной работы;
- результаты самостоятельной работы.

По каждому объекту дается характеристика процедур оценивания в привязке к используемым оценочным средствам.

Промежуточная аттестация студентов. Промежуточная аттестация студентов по дисциплине «Большие данные» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

В зависимости от вида промежуточного контроля по дисциплине и формы его организации могут быть использованы различные критерии оценки знаний, умений и навыков.

Промежуточная аттестация по дисциплине «Методы исследования предметной области» проводится в виде зачета.

Оценочные средства для промежуточной аттестации

Вопросы к зачету

1. Что означает термин «Big Data» в информационных технологиях?
2. Что является основной целью обработки Big Data?
3. Кто и в каком году впервые ввел термин «Big Data»?
4. Какие главные характеристики Big Data?
5. Какие данные занимают больше мировой памяти относительно остальных?
6. Какие понятия содержит в себе принцип трех «V»?
7. С какого года Большие данные изучаются как академический предмет в вузовских программах по науке о данных?
8. Что является примером квази-структурированных данных?
9. Как назывался первый суперкомпьютер, оснащенный вопросно-ответной системой искусственного интеллекта?
10. Чем характеризуются «Большие данные»?
11. Что является главным результатом процесса Business Intelligence?
12. Что означает термин «Business Intelligence» в информационных технологиях?
13. Расшифруйте аббревиатуру OLAP.
14. Что относится к средствам предоставления информации в Business Intelligence?
15. Что относится к средствам интеграции в «Business Intelligence»?
16. Какие цели ставит перед собой Data Science?
17. Что такое жизненный цикл аналитики данных?
18. Дайте определение термину «предиктивное моделирование»?
19. Что такое ETL?
20. Какова роль BI-аналитика в проекте?
21. Что такое Apache Hadoop?
22. В чем преимущества решений на базе Hadoop?
23. Что такое MapReduce?
24. Какими достоинствами и недостатками обладает MapReduce?
25. Какому основному принципу следует HDFS?
26. Какой размер блока по умолчанию в HDFS?
27. Какие функции выполняет NameNode в HDFS?
28. Какой узел отвечает за репликацию данных в Hadoop?
29. Какие компоненты содержит Slave узел в Hadoop?
30. Какие компоненты содержит Master узел в Hadoop?
31. Какие компоненты являются частями HDFS?

32. Какое API было добавлено в Hadoop v2.0?
33. Для чего используется автономный режим Hadoop?
34. Какой режим необходим для того, чтобы на локальной машине использовать Hadoop как кластер, состоящий из одного узла?
35. Что является отличительной особенностью NoSQL?
36. В каком случае стоит применять NoSQL хранилища?
37. Что, согласно теореме CAP, возможно обеспечить в любой реализации распределённых вычислений?
38. Какое свойство означает, что транзакции не нарушают согласованность данных, то есть они переводят базу данных из одного корректного состояния в другое?
39. Какой способ хранения данных используется в MongoDB?
40. Что относится к плюсам репликации?
41. Что относится к преимуществам нереляционных БД?
42. На какие три группы подразделяют пользователей в MongoDB?
43. Что такое шардинг?
44. Какие три свойства фигурируют в определении теоремы CAP?
45. Для чего нужна визуализация?
46. Как называется один из самых популярных языков сценариев?
47. Какие достоинства у Amazon S3?
48. Какие традиционные виды визуализации?
49. Какие отличия и основные возможности у языка R?
50. В чем особенности хранения в Amazon S3?
51. Что такое дедупликация данных?
52. В чем основные задачи визуализации?
53. Какие требования предъявляются к визуализации?
54. Какие типы визуализации можно выделить?
55. Чем анализ больших данных отличается от традиционного анализа?
56. Какие основные типы Data Mining?
57. Какие категории Web Mining можно выделить?
58. В чем основная задача Web Content Mining?
59. В чем основные задачи интеллектуального анализа текстов?

Критерии оценки знаний студентов на зачёте

Студент допускается к сдаче зачёта только при своевременном получении оценки «зачтено» по всем лабораторным работам курса.

Оценка «зачтено» выставляется студенту, который

- выполнил все лабораторные и/или практические работы в должном качестве и в срок;

- прочно усвоил предусмотренный программный материал;
- правильно, аргументировано ответил на все вопросы, с приведением примеров;
- показал глубокие систематизированные знания, владеет приемами рассуждения и сопоставляет материал из разных источников: теорию связывает с практикой, другими темами данного курса, других изучаемых предметов

- без ошибок выполнил практическое задание.

Обязательным условием выставленной оценки является правильная речь в быстром или умеренном темпе.

Дополнительным условием получения оценки «зачтено» могут стать хорошие успехи при выполнении самостоятельной и контрольной работы, систематическая активная работа на семинарских занятиях.

Оценка «не зачтено» выставляется студенту, который

- выполнил все лабораторные и/или практические работы в должном качестве и в срок;
- не справился с 50% вопросов и заданий билета;
- в ответах на другие вопросы допустил существенные ошибки;
- не может ответить на дополнительные вопросы, предложенные преподавателем;
- не может продемонстрировать целостное представление о взаимосвязях, компонентах, этапах развития предмета изучения.

Оценивается качество устной и письменной речи, как и при выставлении положительной оценки.

Оценочные средства для текущего контроля

Текущий контроль предполагает систематическую проверку усвоения учебного материала, сформированности компетенций или их элементов, регулярно осуществляемую на протяжении изучения дисциплины, в соответствии с ее рабочей программой.

Состоит в проверки правильности выполнения заданий по самостоятельной работе. Задание зачтено, если нет ошибок. По текущим ошибкам даются пояснения.

Критерии оценки проектов

- 100-86 баллов выставляется, если обучающийся точно определил содержание и составляющие части задания, умеет аргументировано отвечать на вопросы, связанные с заданием. Продемонстрировано знание и владение

навыками самостоятельной исследовательской работы по теме. Фактических ошибок, связанных с пониманием проблемы, нет.

- 85-76 баллов – работа характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

- 75-61 балл – проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы

- 60-50 баллов – если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы

Шкала оценивания проектов

Менее 60 баллов	Не зачтено
От 61 до 75 баллов	зачтено
От 76 до 85 баллов	зачтено
От 86 до 100 баллов	зачтено

Примерные тесты

1. В принцип «Трёх V» не входит один из перечисленных признаков:	Ответы: а. Veracity (достоверность данных): в настоящее время достоверность имеющихся данных является важнейшим критерием для пользователей. Недостоверная информация приводит к затруднению анализа данных. б. Volume (объем): накопленная база данных представляет собой гигантский объем информации, для которого обработка и хранение традиционными способами являются трудоёмкими процессами. Такой объем нуждается в новых подходах и в более усовершенствованных инструментах. в. Variety (многообразие): данная характеристика означает возможность одновременной обработки
--	---

	структурированной и неструктурированной информации различных форматов. Главным отличием структурированной информации является возможность классификации.
2. К полуструктурированным данным относятся:	<p>Ответы:</p> <p>а. Данные, которые не имеют определённой формы, могут включать в себя видео, аудио файлы, свободный текст, информацию, поступающую из социальных сетей.</p> <p>б. Данные, которые не соответствуют чёткой структуре таблиц и отношений в реляционных базах данных, однако такие данные содержат специальные теги и иные маркёры, позволяющие отделить семантические элементы.</p> <p>в. Данные, определяющие конкретную предметную область, упорядоченные специальным образом и организованные таким образом, чтобы над ними можно было выполнить анализ.</p>

1. Предиктивное моделирование (Predictive Modelling) – это:	<p>а. Процесс создания (или выбора) модели для предсказания вероятности наступления некоторого события.</p> <p>б. Компьютерная техника извлечения знаний, которая использует искусственный интеллект для распознавания образов и выделения значимых закономерностей из данных, находящихся в хранилищах или входных или выходных потоках.</p> <p>в. Методы и инструменты для перевода необработанной информации в осмысленную, удобную форму.</p>
2. Text Mining – это:	<p>Ответы:</p> <p>а. Нетривиальный процесс обнаружения действительно новых, потенциально полезных и понятных шаблонов в неструктурированных текстовых данных.</p> <p>б. Использование методов интеллектуального анализа для автоматического обнаружения веб-документов и услуг, извлечения информации из веб-ресурсов и сервисов.</p> <p>в. Собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.</p>

<p>1. Достоинством модели Map Reduce не является:</p>	<p>Ответы:</p> <ul style="list-style-type: none">а. Автоматическое распараллеливание (функции Map и Reduce могут выполняться параллельно и независимо друг от друга)б. Масштабируемость (данные могут располагаться и обрабатываться на разных серверах).в. Фиксированный алгоритм обработки данных.г. Отказоустойчивость (при отказе сервера функции Map и Reduce запускаются на другом сервере).
<p>1. Ошибка 2-ого рода – это:</p>	<p>Ответы:</p> <ul style="list-style-type: none">а. «Ложная тревога», когда при отсутствии события ошибочно выносится решение о его присутствии.б. «Пропуск цели», когда интересующее событие ошибочно не обнаруживается.