

Министерство науки и высшего образования Российской Федерации  
ФГБОУ ВО «Бурятский государственный университет»  
Институт математики и информатики  
Кафедра информационных технологий

Утверждена на заседании  
Ученого совета ИМИ  
10 сентября 2020 г.  
Протокол №09-20

Рабочая программа дисциплины

## **Машинное обучение**

Направление подготовки  
01.04.01 Математика

Квалификация  
магистр

Форма обучения  
очная

Улан-Удэ  
2021

## Пояснительная записка

### Цели освоения дисциплины

Дать представление о современном инструментарии обработки больших данных.

### Место дисциплины в структуре образовательной программы

Дисциплина относится к вариативной части, является обязательной дисциплиной. Код дисциплины Б1.В.ОД.4

### Планируемые результаты обучения по дисциплине и индикаторы достижения компетенций.

#### В результате освоения дисциплины студент должен:

##### Знать:

Общие принципы работы с большими данными; Основные концепции вычислительных технологий больших данных; Типовые задачи обработки больших данных

##### Уметь:

Осуществлять сбор и хранение больших данных; применять основные концепции вычислительных технологий больших данных; решать типовые задачи обработки больших данных с применением современного инструментария

##### Владеть:

Навыками работы со стеком технологий Hadoop; навыками работы с конкретными инструментами стека технологий Hadoop; навыками решения типовых задач

#### Планируемые результаты освоения образовательной программы:

- ОПК-2 - Способен строить и анализировать математические модели в современном естествознании, технике, экономике и управлении
  - ОПК-2.1 - Подбирает и применяет известные подходы математического моделирования
  - ОПК-2.2 - Подбирает и применяет конкретные методы и средства построения и анализа математических моделей

### Объем дисциплины в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 9 зачетные единицы, 324 часа.

№	Название разделов дисциплины	Лекция	Самостоятельная работа	Лабораторная работа
Семестр 1		22	136	22
1	Стек технологий Hadoop и методы хранения больших данных-1	22	136	22
Семестр 2		12	48	12
1	Стек технологий Hadoop и методы хранения больших данных-2	12	48	12

### Тематическое планирование курса

Стек технологий Hadoop и методы хранения больших данных-1

Семестр 1

#### Сбор и хранение данных

Лекция. 2 ч. модель вычислений Map-Reduce

Лекция. 4(0) ч. Файловая система HDFS

Лекция. 4(0) ч. Посредник YARN и планировщик Oozie

Самостоятельная работа. 40(0) ч. Планировщик Oozie

Лекция. 4(0) ч. NoSQL базы данных

Самостоятельная работа. 30(0) ч. NoSQL база данных Cassandra

Лабораторная работа. 8(0) ч. Импорт данных

Лекция. 4(0) ч. Импорт с помощью Apache Sqoop

Лекция. 4(0) ч. Импорт с помощью Apache Flume

Самостоятельная работа. 34(0) ч. Импорт данных в HDFS

Самостоятельная работа. 32(0) ч. NoSQL база данных HBase

Лабораторная работа. 8(0) ч. Реализация Map-Reduce в Hadoop

Лабораторная работа. 6 ч. Файловая система HDFS

## Стек технологий Hadoop и методы хранения больших данных-2

Семестр 2

### Обработка данных

Лекция. 2 ч. Принципы работы с Apache Hive

Лекция. 2 ч. Принципы работы с Apache Pig

Лекция. 2 ч. HQL (Hive query language)

Лекция. 2 ч. Язык Pig Latin. Структуры данных Tuple и Bag. Базовые функции Pig Latin

Лабораторная работа. 2 ч. Язык запросов Hive query language

Самостоятельная работа. 16 ч. Pig Latin

Самостоятельная работа. 16 ч. Настройка и запуск Hadoop и Pig

Лабораторная работа. 4 ч. Типы данных в Pig Latin

### Иные технологии

Лекция. 2 ч. Vowpal Wabbit

Лабораторная работа. 4 ч. Vowpal Wabbit

Самостоятельная работа. 8 ч. Vowpal Wabbit

Лекция. 2 ч. Фреймворк Caffe и концепция deep learning

Лабораторная работа. 2 ч. Работа с Caffe

Самостоятельная работа. 8 ч. Работа с Caffe

## БРС

Семестр	Контрольные точки	Баллы
1	<b>Текущий контроль</b> в разделе «Стек технологий Hadoop и методы хранения больших данных-1»	
	Разработка проекта	40
	Составление структурно-логической схемы	20
1	<b>Экзамен</b>	
	Разработка проекта	40
Итого за семестр 1: 100		
2	<b>Текущий контроль</b> в разделе «Стек технологий Hadoop и методы хранения больших данных-2»	
	Разработка проекта	40
	Составление структурно-логической схемы	20
2	<b>Экзамен</b>	
	Разработка проекта	40
Итого за семестр 2: 100		

## Учебно-методическое и информационное обеспечение учебного процесса

### Образовательные технологии (в том числе на занятиях, проводимых в интерактивных формах).

При изучении данного курса применяются как традиционные (лекции, практические занятия, экзамен), так и инновационные образовательные технологии, которые реализуются в учебном процессе в активных и интерактивных формах проведения занятий, из которых можно выделить следующие:

- 1) лекция-дискуссия при рассмотрении тем "NoSQL базы данных", "Платформа Kaggle".
- 2) метод группового решения задач при изучении тем "Работа с Caffe", "Работа с Spark MLlib".
- 3) перекрестная самопроверка при изучении тем "Настройка и запуск Hadoop и Pig", "Hadoop Cloudera".

Учебно-методические материалы, в том числе методические указания для обучающихся по освоению дисциплины

К современному специалисту общество предъявляет достаточно широкий перечень требований, среди которых немаловажное значение имеет наличие у выпускников определенных способностей и умения самостоятельно добывать знания из различных источников, систематизировать полученную информацию, давать оценку конкретной ситуации. Формирование такого умения происходит в течение всего периода обучения через участие студентов в лабораторных, практических занятиях, выполнение заданий и тестов, написание курсовых и выпускных квалификационных работ. При этом самостоятельная работа студентов играет решающую роль в ходе всего учебного процесса.

### Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

По данной дисциплине разработано учебно-методическое обеспечение для самостоятельной работы обучающихся и размещено в электронной информационно-образовательной среде университета (личном кабинете студента).

- [CPC\\_mag\\_010401\\_Математика\\_2015\\_Технологии сбора и обработки больших данных\\_05151m.doc](#)

Режим

доступа:

[https://my.bsru.ru/content/file/3/36/363/103424\\_srs\\_mag\\_010401\\_matematika\\_2015\\_tehnologii-sbora-i-obrabotki-bolshih-dannih\\_05151m.doc](https://my.bsru.ru/content/file/3/36/363/103424_srs_mag_010401_matematika_2015_tehnologii-sbora-i-obrabotki-bolshih-dannih_05151m.doc)

### **Учебно-методические материалы, в том числе методические указания для обучающихся по освоению дисциплины**

— На лекционных занятиях студент слушает доклад преподавателя, составляет конспект лекции. Во время лекции студенту рекомендуется делать отметки на полях тетради, касающиеся того теоретического материала, который вызвал затруднения в понимании. После лекции трудности необходимо устранить путем консультации у преподавателя или самостоятельной работы с рекомендованной учебной литературой.

— На лабораторных занятиях студенту предлагается ряд заданий по теме, прослушанной на лекции. По заданию необходимо решить поставленную задачу, пользуясь персональным компьютером и необходимым программным обеспечением. Контроль знаний, умений и навыков студентов осуществляется путём проведения на лабораторных занятиях компьютерных тестов, сдачи проектов. Студенту, выполнившему то или иное задание на низкое количество баллов, по согласованию с преподавателем, необходимо выполнить работу над ошибками для улучшения результатов.

— Предусмотрена самостоятельная работа студентов, в рамках которой студентом должны выполняться работа над проектами, анализ требований, ошибок, перспектив и т.п.. Теоретический материал студентом должен быть проработан с использованием конспектов лекций (если в плане предусмотрены лекции) или рекомендуемой учебной литературы.

Советы по планированию и организации времени, необходимого для изучения дисциплины.

Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины:

В день после лекционного занятия в течение 1 астрономического часа изучить лекцию, выявить моменты, где имеются вопросы и недостаточное понимание. По этим вопросам изучить рекомендованную литературу, устранить непонимание. В день перед проведением практического занятия в течение 30 минут повторить пройденный материал, предшествующий теме занятия.

При изучении дисциплины очень полезно самостоятельно изучать материал, который еще не прочитан на лекции, не применялся на лабораторном или практическом занятии. Тогда предстоящее занятие будет гораздо понятнее.

Рекомендации по работе с литературой. Теоретический материал курса становится более понятным, когда дополнительно к прослушиванию лекций, изучению конспектов, изучаются и книги по изучаемой дисциплине. Литературу по дисциплине рекомендуется изучать в библиотеке. Полезно использовать несколько учебников по дисциплине. Рекомендуется, кроме «заучивания» материала, добиться состояния понимания изучаемой темы дисциплины. С этой целью рекомендуется после изучения очередного параграфа выполнить несколько простых упражнений на данную тему.

Советы по подготовке к экзамену:

1) Ознакомиться с процедурой проведения экзамена, ознакомиться со списком вопросов для подготовки к экзамену. Повторить весь материал пройденных лекций (если в плане были лекции), проработать конспекты, рекомендованную учебную литературу. Составить конспекты ответов на экзаменационные вопросы.

2) Собрать итоговые выполненные лабораторные и проекты (если в плане были лабораторные и проекты), если произведено их улучшение согласно замечаниям преподавателя. В рамках экзамена/зачета возможен пересмотр преподавателем баллов (по усмотрению преподавателя, с учетом общей работы студента во время курса), выставленных ранее за выполненные задания, при условии получения улучшения результатов

### **Оценочные средства**

По данной дисциплине разработаны оценочные средства, критерии их оценивания, а также методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

- [CPC\\_mag\\_010401\\_Математика\\_2015\\_Технологии сбора и обработки больших данных\\_05151m.doc](#)
- [ФОС\\_mag\\_010401\\_Математика\\_2015\\_Технологии сбора и обработки больших данных\\_05151m.docx](#)

### **Список литературы**

Перечень основной и дополнительной литературы, необходимой для освоения дисциплины.

Основная

1. [ОСНОВЫ ИСПОЛЬЗОВАНИЯ И ПРОЕКТИРОВАНИЯ БАЗ ДАННЫХ](#): Учебник/Илюшечкин В.М.. —М.: Издательство Юрайт, 2016. —213 с.  
Режим доступа: <http://www.biblio-online.ru/book/1C650A7F-DC7D-4834-998E-42D06FC8EF33>

2. [ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ](#): Учебник и практикум/Миркин Б.Г.. —М.: Издательство Юрайт, 2016. —174 с.  
Режим доступа: <http://www.biblio-online.ru/book/E486A3B0-1AE9-4179-8D48-FA24B626C7C9>
3. [АНАЛИЗ ДАННЫХ](#): Учебник/Мхитарян В.С. - Отв. ред.. —М.: Издательство Юрайт, 2016. —490 с.  
Режим доступа: <http://www.biblio-online.ru/book/AF1D197F-1759-422E-9593-8B43E2D1093B>

Дополнительная

1. [Применение искусственных нейронных сетей и системы остаточных классов в криптографии](#)/[Н.И. Червяков и др.]. —Москва: Физматлит, 2012. —279 с.  
Режим доступа: [http://e.lanbook.com/books/element.php?pl1\\_cid=25&pl1\\_id=5300](http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=5300)
2. [Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных](#): научно-популярная литература/П. Флах ; пер. с англ. А. А. Слинкин. —Москва: ДМК Пресс, 2015. —400 с.  
Режим доступа: [http://e.lanbook.com/books/element.php?pl1\\_id=69955](http://e.lanbook.com/books/element.php?pl1_id=69955)
3. [Scala для нетерпеливых](#)/К. Хостманн. —Москва: ДМК Пресс, 2013. —408 с.  
Режим доступа: [http://e.lanbook.com/books/element.php?pl1\\_id=66473](http://e.lanbook.com/books/element.php?pl1_id=66473)
4. [Python и анализ данных](#): научно-популярная литература/У. Маккинли ; [пер. с англ. А. А. Слинкин]. —Москва: ДМК Пресс, 2015. —482 с.  
Режим доступа: [http://e.lanbook.com/books/element.php?pl1\\_id=73074](http://e.lanbook.com/books/element.php?pl1_id=73074)
5. [Python на практике](#)/Саммерфилд М.. —Москва: ДМК Пресс, 2014  
Режим доступа: [http://e.lanbook.com/books/element.php?pl1\\_id=66480](http://e.lanbook.com/books/element.php?pl1_id=66480)

**Перечень ресурсов информационно-коммуникационной сети «Интернет», необходимых для освоения дисциплины**

<https://www.ibm.com/developerworks/ru/library/os-log-process-hadoop/>  
<https://www.kaggle.com/c/digit-recognizer>

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

<http://hadoop.apache.org/>

**Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Учебная аудитория для проведения лекционных занятий (1205, 1209, 1208, 1211); Учебная аудитория для проведения занятий лабораторного типа с доступом в Интернет (1313, 1312, 1316); Помещение для самостоятельной работы с доступом в Интернет (1312, 1313, 1316);

Учебная аудитория для проведения индивидуальных и групповых консультаций (1313, 1312, 1316); Учебная аудитория для проведения текущей и промежуточной аттестации (1312, 1316, 1313); Требуемый перечень программного обеспечения: ОС Windows/Ubuntu; Интегрированная среда разработки Code::Blocks 13.12 и выше.

Автор: Цыбиков Анатолий Сергеевич

Рабочая программа обсуждена на заседании кафедры информационных технологий от 09 сентября 2020 г.  
Протокол №09-20-2.

Рабочая программа одобрена на заседании Учебно-методической комиссии ИМИ от 10 сентября 2020 г.  
Протокол №1.