

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Беспалов Владимир Александрович

Должность: Ректор МИЭТ

Дата подписания: 01.09.2025 12:25:22

Уникальный программный ключ:

ef5a4fe6ed0ffdf3f1a49d6ad1b49464dc1bf7354f736d76e016bea882b88b02

Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет

«Московский институт электронной техники»

УТВЕРЖДАЮ

Проректор по учебной работе

И.Г. Игнатова

«21» 06 2021 г.



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Наука о данных и аналитика больших данных»

Направление подготовки – 09.03.03 «Прикладная информатика»

Направленность (профиль) – «Системы корпоративного управления»

Москва 2020

1. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

Дисциплина участвует в формировании следующих компетенций образовательной программы:

Компетенция	Подкомпетенция, формируемая в дисциплине	Индикаторы достижения компетенций
ОПК-2 способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК – 2.НДАБД способен изучать, осваивать, разрабатывать и внедрять новые технологии обработки и анализа больших массивов данных	Знания стандартов обработки и анализа больших данных, и требований, связанных с созданием и использованием SQL и NoSQL систем хранения и обработки данных Умения использовать современные инструментальные и вычислительные средства (в соответствии с профилем подготовки) Опыт визуализации результатов аналитики при помощи языка R

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина входит в обязательную часть Блока 1 «Дисциплины (модули)» образовательной программы.

Входные требования: для изучения дисциплины необходимы базовые знания основ программирования (алгоритмы и структуры данных, ООП, паттерны проектирования); понимание принципов проектирования СУБД и знание SQL; знания одного или нескольких языков программирования: Java, Perl, Python и т.д.; понимания основ статистической обработки данных.

3. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Курс	Семестр	Общая трудоёмкость (ЗЕ)	Общая трудоёмкость (часы)	Контактная работа			Самостоятельная работа (часы)	Промежуточная аттестация
				Лекции (часы)	Лабораторные работы (часы)	Практические занятия (часы)		
4	7	2	72	16	16	-	40	За

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

№ и наименование модуля	Контактная работа			Самостоятельная работа	Формы текущего контроля
	Лекции (часы)	Лабораторные работы (часы)	Практические занятия (часы)		
1. Наука о данных и аналитика больших данных	16	16	-	40	Итоговый контрольный тест

4.1. Лекционные занятия

№ модуля дисциплины	№ лекции	Объем занятий (часы)	Краткое содержание
1	1	2	Введение в большие данные. Определение, терминология, история и причины появления Big Data, характеристики, источники больших данных, задачи, решаемые с помощью больших данных
	2	2	Жизненный цикл аналитики данных. Business Intelligence. ETL-процесс. Технология OLAP. Предиктивное моделирование и Data Mining
	3	2	Высокопроизводительные вычисления. История Hadoop и Map Reduce
	4	2	Масштабирование и многоуровневое хранение данных. Основы NoSQL. MongoDB
	5	2	Визуализация данных и результатов анализа. Типы и задачи визуализации. Дашборды. Основные возможности языка R
	6	2	Сложные методы аналитики. Машинное обучение. Кластерный анализ.
	7	2	Анализ текста. Особенности анализа текста, варианты использования. Полнотекстовый поиск
	8	2	Обобщение изученного материала

4.2. Практические занятия

Не предусмотрены

4.3. Лабораторные работы

№ модуля дисциплины	№ лабораторной работы	Объем занятий (часы)	Наименование работы
1	1	2	Выбор предметной области. Формирование набора данных

№ модуля дисциплины	№ лабораторной работы	Объем занятий (часы)	Наименование работы
	2	2	Архитектура проектируемой системы
	3	2	Хранение данных
	4	2	Обработка данных
	5	2	Визуализация результатов
	6	2	Установка и работа с Apache Web-server. Разработка MAP-REDUCE задачи.
	7	2	Создание базы данных в HBase
	8	2	Установка и работа с библиотеками Apache. Создание и операции над RDD.

4.4. Самостоятельная работа студентов

№ модуля дисциплины	Объем занятий (часы)	Вид СРС
1	34	Выполнение практико-ориентированных заданий по темам лабораторных работ. Подготовка и выполнение тестов самоконтроля. Просмотр видеолекций
	6	Подготовка к итоговому тестированию

4.5. Примерная тематика курсовых работ (проектов)

Не предусмотрены

5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Учебно-методическое обеспечение для самостоятельной работы студентов в составе УМК дисциплины (ОРИОКС, <http://orioks.miet.ru/>):

1. Наука о данных и аналитика больших данных

✓ материалы для подготовки к лабораторным работам: лабораторный практикум с описанием теоретических сведений, примеров выполнения задания, порядка выполнения задания.

✓ материалы для подготовки к тестам: материалы лекций и видеоматериалы, теоретический материал практических заданий и лабораторных работ.

6. ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ

Литература

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Электрон. дан. — Москва: ДМК Пресс, 2015. — 400 с. — URL: <https://e.lanbook.com/book/69955> (дата обращения: 10.11.2020)
2. Чак, Л. Hadoop в действии / Л. Чак. — Электрон. дан. — Москва: ДМК Пресс, 2012. — 424 с. — URL: <https://e.lanbook.com/book/39997> (дата обращения: 10.11.2020)
3. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д.У. Джефффри; пер. с англ. Слинкин А.А.. — Электрон. дан. — Москва: ДМК Пресс, 2016. — 498 с. — URL: <https://e.lanbook.com/book/93571> (дата обращения: 10.11.2020)
4. От хранения данных к управлению информацией = Information Storage and Management : Storing, Managing, and Protecting Digital Information / Ed.G. Somasundaram, Alok Shrivastava; EMC Education Services. - СПб. : Питер, 2010. - 524 с. - ISBN 978-5-4237-0008-9
5. Храмов, Д.А. Сбор данных в Интернете на языке R / Д.А. Храмов. — Электрон. дан. — Москва: ДМК Пресс, 2017. — 280 с. — URL: <https://e.lanbook.com/book/97350> (дата обращения: 10.11.2020)

7. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ БАЗ ДАННЫХ, ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ

1. Apache HBase : [сайт].- На англ. языке. - URL: <https://hbase.apache.org/> (дата обращения: 10.11.2020)
2. Apache Zookeeper : [сайт].- На англ. языке. - URL: <https://zookeeper.apache.org> (дата обращения: 10.11.2020)
3. Hue : [сайт].- На англ. языке. - URL: <http://gethue.com> (дата обращения: 10.11.2020)
4. R language fundamentals. Data frames. Steven Buechler : [сайт].- На англ. языке. - URL: <https://pdfslide.net/documents/r-language-fundamentals-data-frames.html> (дата обращения: 10.11.2020)
5. Python Pandas : [сайт].- На англ. языке. - URL: <http://pandas.pydata.org/> (дата обращения: 10.11.2020)
6. Apache Parquet : [сайт].- На англ. языке. - URL: <https://parquet.apache.org> <https://zookeeper.apache.org> (дата обращения: 10.11.2020)

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В ходе реализации обучения используется смешанное обучение, сочетающее традиционные формы аудиторных занятий (в компьютерном классе) и взаимодействие в электронной образовательной среде.

Освоение образовательной программы обеспечивается ресурсами электронной информационно-образовательной среды ОРИОКС(<http://orioks.miet.ru>).

Для взаимодействия студентов с преподавателем используются сервисы обратной связи: разделы ОРИОКС «Новости», «Домашние задания»; электронная почта, каналы Discord, Zoom.

В процессе обучения при проведении занятий и для самостоятельной работы используются **внутренние электронные ресурсы** в формах видеолекций и тестирования в ОРИОКС.

При проведении занятий и для самостоятельной работы используются **внешние электронные ресурсы** в формах *электронных компонентов сервисов*:

1. Официальный сайт Apache Hadoop, предоставляющий возможность скачать любую версию фреймворка [Электронный ресурс]. - Режим доступа: <http://www.apache.org/dyn/closer.cgi/hadoop/common/>
2. Пользовательское руководство по HDFS от Apache [Электронный ресурс]. - Режим доступа: <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html>
3. Документация Hadoop по TaskTracker [Электронный ресурс]. - Режим доступа: <https://wiki.apache.org/hadoop/TaskTracker>
4. Документация Hadoop по DataNode. [Электронный ресурс]. Режим доступа: <https://wiki.apache.org/hadoop/DataNode>
5. Документация Hadoop по JobTracker. [Электронный ресурс]. Режим доступа: <https://wiki.apache.org/hadoop/JobTracker>
6. Документация Hadoop по NameNode. [Электронный ресурс]. Режим доступа: <https://wiki.apache.org/hadoop/NameNode>
7. Документация Oracle по jps – команде, отображающей Java-процессы. [Электронный ресурс]. Режим доступа: <http://docs.oracle.com/javase/7/docs/technotes/tools/share/jps.html>
8. Документация Hadoop по NodeManager. [Электронный ресурс]. Режим доступа: <http://hortonworks.com/blog/apache-hadoop-yarn-nodemanager/>
9. Документация Hadoop по ResourceManager. [Электронный ресурс]. Режим доступа: <http://hortonworks.com/blog/apache-hadoop-yarn-resourcemanager/>
10. Список команд HDFS. [Электронный ресурс]. Режим доступа: <http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html>
11. Список примеров MapReduce задач, доступных в Hadoop. [Электронный ресурс]. Режим доступа: <http://www.informit.com/articles/article.aspx?p=2190194&seqNum=3>
12. Документация по Spark Context. [Электронный ресурс]. Режим доступа: <https://spark.apache.org/docs/1.4.0/api/java/org/apache/spark/SparkContext.html>
13. Документация Spark по RDD. [Электронный ресурс]. Режим доступа: <http://spark.apache.org/docs/1.2.1/programming-guide.html#resilient-distributed-datasets-rdds>
14. Документация по классу Tuple. [Электронный ресурс]. Режим доступа: <http://www.scala-lang.org/api/2.9.3/scala/Tuple2.html>
15. Документация Spark по DataFrame. [Электронный ресурс]. Режим доступа: <https://spark.apache.org/docs/1.6.0/sql-programming-guide.html#dataframes>
16. Документация по SQLContext. Официальный сайт. [Электронный ресурс]. Режим доступа: <https://spark.apache.org/docs/1.6.1/sql-programming-guide.html#starting-point-sqlcontext>
17. Документация по набору операций для DataFrame. [Электронный ресурс]. Режим доступа: <https://spark.apache.org/docs/1.6.0/sql-programming-guide.html#dataframe-operations>
18. Шитиков В. К., Мاستицкий С. Э. Классификация, регрессия, алгоритмы Data Mining с использованием R : Электронная книга. – 2017. – Режим доступа: <https://github.com/ranalytics/data-mining> (дата обращения 09.04.2019)

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Наименование учебных аудиторий и помещений для самостоятельной работы	Оснащенность учебных аудиторий и помещений для самостоятельной работы	Перечень программного обеспечения
Компьютерный класс	Компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в ОРИОКС	Azure (Windows), Microsoft Office Pro, 7z, Acrobat Reader DC, Доступ к ПО через удаленный рабочий стол skylab.sipc.miet.ru: Azure (Microsoft Visual Studio)
Помещение для самостоятельной работы	Компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в ОРИОКС	ОС Microsoft Windows, Microsoft Office Professional Plus, Google Chrome, Acrobat reader DC, Apache Web-server, Microsoft Visual Studio

10. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕРКИ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

ФОС по подкомпетенции ОПК-2.НДАБД «Способен изучать, осваивать, разрабатывать и внедрять новые технологии обработки и анализа больших массивов данных».

Фонды оценочных средств представлены отдельными документами и размещены в составе УМК дисциплины электронной информационной образовательной среды ОРИОКС// URL: <http://www.orioks.miet.ru/>.

11. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

11.1. Особенности организации процесса обучения

В дисциплине предусмотрены следующие виды занятий: лекции, лабораторные работы и самостоятельная работа. Форма промежуточного контроля – зачет.

На лекционных занятиях преподаватель объясняет тему, в процессе объяснения студенты задают уточняющие вопросы, ведется дискуссия.

На лабораторных работах выполняется индивидуальная работа в выбранной области анализа. В процессе выполнения этого задания студенты закрепляют навыки, полученные на лекционных занятиях. Также студенты выполняют задания лабораторного практикума. После каждой пройденной темы студенты выполняют тесты самоконтроля.

В процессе изучения курса преподавателем проводятся консультационные занятия. На консультациях студентам даются пояснения по трудноусваиваемым разделам дисциплины. Допускается задать вопрос преподавателю и по электронной почте.

11.2. Система контроля и оценивания

Для оценки успеваемости студентов по дисциплине используется балльная накопительная система.

Баллами оцениваются: выполнение каждого контрольного мероприятия в семестре (в сумме до 70 баллов), сдача зачета (в сумме до 30 баллов) и активность в семестре, как бонусное контрольное мероприятие (в сумме до 10 баллов). По сумме баллов выставляется итоговая оценка по предмету. Структура и график контрольных мероприятий приведены в журнале успеваемости на ОРИОКС, <http://orioks.miet.ru/>.

Мониторинг успеваемости студентов проводится в течение семестра трижды: по итогам 1-8 учебных недель, 9 – 12 учебных недель, 13 – 18 учебных недель.

РАЗРАБОТЧИК:

Ассистент

/ А.А. Рубцова /

К.т.н, старший научный сотрудник, доцент

/ Г.Н. Брусникин /

Рабочая программа дисциплины «Наука о данных и аналитика больших данных» по направлению подготовки 09.03.03 «Прикладная информатика», направленности (профилю) «Системы корпоративного управления» разработана в институте СПИНТех и утверждена на заседании института 15 июня 2021 года, протокол № 6.

Директор института СПИНТех

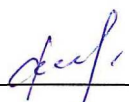
 / Л.Г. Гагарина/

ЛИСТ СОГЛАСОВАНИЯ

Программа согласована с Центром подготовки к аккредитации и независимой оценке качества

Начальник АНОК  / И.М. Никулина /

Программа согласована с библиотекой МИЭТ

Директор библиотеки  / Т.П. Филиппова /