

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Беспалов Владимир Александрович

Должность: Ректор МИЭТ

Дата подписания: 01.09.2020 14:25:33

Уникальный программный ключ:

ef5a4fe6ed0ffdf3f1a49d6ad1b49464dc1bf7354f736a76c8f8baa882b8d602

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

« 9 »

2020 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Большие данные»

Направление подготовки - 09.04.03 «Прикладная информатика»

Направленность (профиль) - «Системы корпоративного управления для инновационных отраслей»

1. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

Дисциплина участвует в формировании следующих компетенций образовательной программы:

Компетенция	Подкомпетенции, формируемые в дисциплине	Индикаторы достижения компетенций
ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.БД способен использовать методы обработки и визуализации больших данных	Знания: методов обработки и представления больших данных; Умения: использовать методы обработки (кластеризации) больших данных; Опыт применения программного обеспечения для обработки больших данных

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина входит в обязательную часть Блока 1 «Дисциплины (модули)» образовательной программы, изучается на 1 курсе во 2 семестре (очная форма обучения).

Входные требования: сформированность компетенций, определяющих готовность работать с базами данных и разрабатывать алгоритмы 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)		
1	2	4	144	16	16	-	76	Экз (36)

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

№ и наименование модуля	Контактная работа			Самостоятельная работа	Формы текущего контроля
	Лекции (часы)	Лабораторные работы (часы)	Практические занятия (часы)		
1. Обработка данных	4	4	-	18	Контроль выполнения и защита Лабораторной работы 1
2. Организация распределенных вычислений	4	4	-	10	Контроль выполнения и защита Лабораторной работы 2
					Тестирование
3. Кластерный анализ больших данных	4	4	-	20	Контроль выполнения и защита Лабораторной работы 3
4. Выявление зависимостей в больших данных с помощью аппарата искусственных нейронных сетей	4	4	-	28	Контроль выполнения и защита Лабораторной работы 4
					Контрольная работа Контроль выполнения самостоятельной работы

4.1. Лекционные занятия

№ модуля дисциплины	№ лекции	Объем занятий (часы)	Краткое содержание
1	1	2	Способы организации хранения больших данных.
	2	2	Накопление и начальная разметка данных
2	3	2	Технология Hadoop.
	4	2	Технология MapReduce.
3	5	2	Способы кластеризации.
	6	2	Критерии оценки проведенной кластеризации.
4	7	2	Машинное и глубинное обучение.
	8	2	Прогностическая функция «больших данных». Контрольная работа.

4.2. Практические занятия

Не предусмотрены

4.3. Лабораторные работы

№ модуля дисциплины	№ лабораторного занятия	Объем занятий (часы)	Наименование занятия
1	1	4	Начальная обработка больших данных, введение метрик оценки.
2	2	4	Организация распределенных вычислений.
3	3	4	Кластеризация больших данных.
4	4	4	Разработка ИНС для анализа больших данных и ее верификация.

4.4. Самостоятельная работа студентов

№ модуля дисциплины	Объем занятий (часы)	Вид СРС
1	18	Обработка набора данных по продаже подержанных автомобилей в Германии за 2016 год, представленного по адресу https://www.kaggle.com/orgesleka/used-cars-database/data . В рамках обработки должна быть проведена проверка на целостность представленного набора данных, заполнение пропусков в данных одним из рассмотренных на лекции способов и введение метрики для дальнейшей работы. Подготовка к выполнению и защите лабораторной работы.
2	10	Сегментация полученного в результате первой работы набора данных для обработки с помощью технологий Hadoop и MapReduce. Подготовка к выполнению и защите лабораторной работы.
3	20	Проведение кластеризации полученных сегментов набора данных методом случайного леса. Выявление правил распределения записей по кластерам, характерных для каждого сегмента, и их экстраполяция на набор данных в целом. Подготовка к выполнению и защите лабораторной работы.
4	28	Разбиение набора данных на тренировочную и тестовую выборку. Создание нейронной сети и ее обучение на тренировочной выборке для

№ модуля дисциплины	Объем занятий (часы)	Вид СРС
		предсказания цены автомобиля и его кластера, который был присвоен ему в предыдущей работе. Оценка качества обучения нейронной сети минимум двумя способами из перечисленных: матрица неточностей, F-мера или ROC-кривая.

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 (<http://orioks.miet.ru/>):

Модуль 1 «Обработка данных»

- ✓ Теоретические сведения
- ✓ Задания на лабораторные работы
- ✓ Методические указания по выполнению СРС
- ✓ Примеры решения задач
- ✓ Задания на самостоятельную работу

Модуль 2 «Организация распределенных вычислений»

- ✓ Теоретические сведения
- ✓ Задания на лабораторные работы
- ✓ Методические указания по выполнению СРС
- ✓ Примеры решения задач
- ✓ Задания на самостоятельную работу

Модуль 3 «Кластерный анализ больших данных»

- ✓ Теоретические сведения
- ✓ Задания на лабораторные работы
- ✓ Методические указания по выполнению СРС
- ✓ Примеры решения задач
- ✓ Задания на самостоятельную работу

Модуль 4 «Выявление зависимостей в больших данных с помощью аппарата искусственных нейронных сетей»

- ✓ Теоретические сведения
- ✓ Задания на лабораторные работы

- ✓ Методические указания по выполнению СРС
- ✓ Примеры решения задач
- ✓ Задания на самостоятельную работу

6. ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ

Литература

1. Ингерсолл, Г. С. Обработка неструктурированных текстов. Поиск, организация и манипулирование / Г. С. Ингерсолл, Т. С. Мортон, Э. Л. Фэррис. — Москва : ДМК Пресс, 2015. — 414 с. — ISBN 978-5-97060-144-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/73069> (дата обращения: 09.02.2021). — Режим доступа: для авториз. пользователей.
2. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения : руководство / С. Рашка ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2017. — 418 с. — ISBN 978-5-97060-409-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/100905> (дата обращения: 09.02.2021). — Режим доступа: для авториз. пользователей.
3. Изучаем Spark: молниеносный анализ данных / Х. Карау, Э. Конвински, П. Венделл, М. Захария. — Москва : ДМК Пресс, 2015. — 304 с. — ISBN 978-5-97060-323-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/90118> (дата обращения: 09.02.2021). — Режим доступа: для авториз. пользователей.
4. Григорьев, Ю. А. Реляционные базы данных и системы NoSQL : учебное пособие / Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова. — Благовещенск : АмГУ, 2018. — 424 с. — ISBN 978-5-93493-308-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/156492> (дата обращения: 09.02.2021). — Режим доступа: для авториз. пользователей.

Периодические издания

1. Информатика и ее применение : Ежеквартальный журнал / Российская академия наук, Федеральный исследовательский центр «Информатика и управление» Российской академии наук. - М. : ТОРУС ПИРЕСС, 2007-. - URL : <http://www.ipiran.ru/journal/issues/> (дата обращения: 22.10.2020)
2. Supercomputing frontiers and innovations : an international open access journal. – Режим доступа: <https://superfri.org/superfri/index> (дата обращения: 19.11.2020)
3. Программные системы : теория и приложения : Электронный научный журнал / Переславль-Залесский : Ин-т программных систем им. А.К. Айламазяна РАН, 2010-. — URL: <http://psta.psiras.ru/archives/archives.html>
4. Программирование / Ин-т системного программирования РАН. - М. : ИКЦ Академкнига, 1975-. - URL: <http://elibrary.ru/contents.asp?titleid=7966> (дата обращения: 01.11.2020). - Переводная версия PROGRAMMING AND COMPUTER SOFTWARE

(составной журнал) <https://link.springer.com/journal/11086> (дата обращения: 19.11.2020)

5. Естественные и технические науки / Издательство "Спутник+". - Сайт журнала <http://www.etn.sc-site.ru/>. Сайт издательства <http://www.sputnikplus.ru/> (дата обращения: 19.11.2020).
6. Компьютер пресс / ООО КомпьютерПресс. - М., 1989 -. - Режим доступа: <http://www.compress.ru> (дата обращения: 19.11.2020)

7. ПЕРЕЧЕНЬ ПРОФЕССИОНАЛЬНЫХ БАЗ ДАННЫХ, ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ

1. ЭБС издательства Лань - <http://e.lanbook.com/> (дата обращения: 01.11.2020).
2. Научная электронная библиотека eLIBRARY.RU URL: <http://elibrary.ru/> (дата обращения: 01.11.2020).
3. Единое окно доступа к информационным ресурсам URL: <http://window.edu.ru/catalog/> (дата обращения: 01.11.2020).

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Может реализовываться с применением электронного обучения и дистанционных образовательных технологий.

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

Для взаимодействия студентов с преподавателем используются сервисы обратной связи: раздел ОРИОКС «Домашние задания», электронная почта, Skype.

В процессе обучения при проведении занятий и для самостоятельной работы используются **внутренние электронные ресурсы**: шаблоны и примеры оформления выполненной работы, разъясняющий суть работы видеоролик, требования к выполнению и оформлению результата.

При проведении занятий и для самостоятельной работы используются внешние электронные ресурсы:

1. Введение в обработку Больших Данных [GeekBrains] – канал YouTube «GeekBrains» - URL: https://www.youtube.com/watch?v=JLNBcSL0HLM&ab_channel=GeekBrains (Дата обращения: 19.11.2020)
2. BigData: хранение и анализ больших данных – канал YouTube «Kharkiv IT Cluster» - URL: https://www.youtube.com/watch?v=SzNr5LXvAv0&ab_channel=KharkivITCluster (Дата обращения: 19.11.2020)
3. Hadoop. Введение в Big Data и MapReduce – канал YouTube «Технострим Mail.Ru Group» - URL: https://www.youtube.com/watch?v=8tzCrau5JuE&ab_channel=ТехностримMail.RuGroup (Дата обращения: 19.11.2020)

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень программного обеспечения
Учебная аудитория	Аудитория с комплектом мультимедийного оборудования	ОС Microsoft Windows, Microsoft Office Professional Plus, Google Chrome, Acrobat reader DC
Компьютерный класс	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, JetBrains PyCharm
Помещение для самостоятельной работы обучающихся	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, JetBrains PyCharm

10. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕРКИ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

ФОС по компетенции/подкомпетенции ОПК-1.БД «Способен использовать методы обработки и визуализации больших 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 – экзамен.

Первым модулем является модуль «Обработка данных», посвященный проблемам организации хранения и алгоритмам базовой разметки больших данных. Во втором модуле происходит изучение технологий Hadoop и MapReduce. При изучении третьего модуля даются общие сведения об алгоритмах кластеризации и оценке результата разбития на кластеры. Четвертый модуль посвящен организации и обучению нейронных сетей для их последующего использования в предикативном ключе.

В лабораторных работах наибольшее внимание уделено соотнесению полученных результатов с логическими выводами для обеспечения наибольшей достоверности.

Выполнение СРС позволяет вовлечь студентов в решение практически значимых задач.

11.2. Система контроля и оценивания

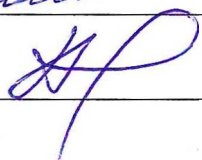
Для оценки успеваемости студентов по дисциплине используется балльная накопительная система.

Баллами оцениваются: выполнение каждого контрольного мероприятия в семестре (в сумме до 60 баллов), активность в семестре (в сумме до 15 баллов) и сдача экзамена (до 25 баллов). По сумме баллов выставляется итоговая оценка по предмету. Структура и график контрольных мероприятий приведены в журнале успеваемости на ОРИОКС (<http://orioks.miet.ru/>).

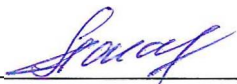
Мониторинг успеваемости студентов проводится в течение семестра трижды: по итогам 1-8 учебных недель, 9 – 12 учебных недель, 13 – 18 учебных недель.

РАЗРАБОТЧИК:

Старший преподаватель СПИНТех  / И.О. Гайдук /

Профессор института СПИНТех, д.т.н.  / Е.М. Портнов/

Рабочая программа дисциплины «Большие данные» по направлению подготовки 09.04.03 «Прикладная информатика», направленности (профилю) «Системы корпоративного управления для инновационных отраслей» разработана в институте СПИНТех и утверждена на заседании института 24 ноября 2020 года, протокол № 3

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 /