

Департамент образования и науки города Москвы
Государственное бюджетное профессиональное образовательное учреждение города Москвы
«Московский колледж управления, гостиничного бизнеса и информационных технологий
«Царицыно»

Принята на заседании
методического совета колледжа
от «29» сентября 2020 г.
Протокол № 1

Утверждаю



Директор колледжа
И.В. Седова

Дополнительная общеобразовательная
общеразвивающая программа
технической направленности

«Анализ данных. Большие данные»

Уровень программы: ознакомительный
Возраст обучающихся: 14-18 лет
Срок реализации: 1 год
Количество часов по программе: 99 часов

Автор-составитель программы:
Соколова А.А., преподаватель ГБПОУ Колледж «Царицыно»

Москва
2020

СОДЕРЖАНИЕ

	стр.
1. Пояснительная записка	3
2. Учебно-тематический план	10
3. Содержание	14
4. Методическое обеспечение	16
5. Материально-техническое обеспечение	16
6. Приложение	19

1. Пояснительная записка

Данные — это конкретные, измеряемые величины, цифры. В мире природы данные — это температура или количество осадков. В мире бизнеса — сумма платежа, количество транзакций, остаток на складе, в мире пользователей — возраст, пол, ежемесячный доход, количество друзей в социальных сетях. Данные — измеряемые и объективные.

Представим, что предприниматель хочет узнать свой доход. Для этого он должен оперировать данными: посчитать количество платежей и умножить на сумму каждого платежа. Если он поленится считать и просто прикинет доход на глаз, то данных не получится, будут только предположения. Предположения — неизмеряемые и субъективные.

Человечество окружает очень много данных, целый океан цифр. Это и есть «большие данные». На самом деле термин big data правильней переводить как «много информации», однако у нас уже есть устоявшийся термин. Большие так большие.

Информации может быть действительно очень много. Например, «Шевроле» устанавливает датчики в обычные легковые автомобили. В 2016 году они накопили и передали в компанию более 4200 терабит информации — хватит, чтобы полностью забить 14 тысяч ноутбуков. Это сотни видов данных: температура масла в двигателе и воздуха в салоне, скорость движения, давление в шинах, работа подвески. Это типичные «большие данные». «Шевроле» использует их, чтобы совершенствовать будущие модели автомобилей.

Почему «большие данные» стали так популярны? Данные полезны клиентам. С помощью «больших данных» реклама становится точнее. Такая реклама нравится клиентам: она не раздражает и рекламирует действительно полезные товары. Стоит зайти в интернет-магазин обуви, как обувь начнет преследовать в рекламе на всех сайтах. Это не меняется, даже если купишь ботинки. Глупое дело: если я купил ботинки, зачем мне дальше их показывать? Вряд ли я куплю еще одни. С «большими данными» магазин узнает, что вы купили обувь. Вместо ботинок вам предложат водоотталкивающий крем — при этом сообщат, что через пару дней будут недельные дожди. А вместе с ним предложат плащ по распродаже и зонтик. Вы же давно не покупали плащей. С «большими данными» реклама перестанет быть инструментом тупого впаривания товаров, а превращается в полезного советчика. Данные помогают компаниям конкурировать. Раньше маркетологи опирались на «маленькие данные»: пол, возраст, место жительства потенциальных клиентов. Со временем такие данные появились у всех конкурентов в отрасли. Выделиться было трудно, реклама стала массовой и раздражающей. С «большими данными» компании научились конкурировать лучше. Они

больше не тратят огромные маркетинговые бюджеты на «ковровые бомбардировки» рынка, а точно вычисляют своего клиента по множеству параметров.

Содержание программы направлено на формирование универсальных учебных действий, обеспечивающих развитие познавательных и коммуникативных навыков. Обучающиеся, включаются в проектную и исследовательскую деятельность, основу которой составляют такие учебные действия, как умение видеть закономерности в данных, строить на этих закономерностях модели, а также проводить валидацию и последующую доработку модели. Обучающиеся включаются в коммуникативную учебную деятельность, где преобладают такие ее виды, как умение полно и точно выражать свои мысли, аргументировать свою точку зрения, работать в группе, представлять и сообщать информацию в устной и письменной форме, вступать в диалог. Решение практических задач аналитики данных позволит ученикам освоить умение применять разработанные решения на практике, тем самым размышляя не только в контексте теоретической модели, но и охватывая предметную область.

Программа курса «Анализ данных. Большие данные» предусматривает реализацию следующих принципов:

- Придать развитию знаний динамичный характер: уметь работать с технологиями информационного поиска, а также уметь анализировать большие объёмы данных.
- Сконцентрировать учебный материал, укрупнив комплектные единицы знаний, что создает дидактические условия для развития системного мышления у учащихся: освободить учебный материал от деталей, имеющих специальное значение, но излишних для общего образования, группируя при этом частные понятия, необходимые для общего образования, вокруг ведущих понятий.
- Формировать у обучающихся аналитическое мышление, сочетая его с активной познавательной и исследовательской деятельностью обучающихся.
- Учитывать возрастные, индивидуальные особенности и возможности обучающихся, предлагая им задания по выбору, а также возможность самостоятельно собирать данные и разрабатывать модели на них.

Данная программа носит пропедевтический характер и активизацию воспитательной деятельности. Обучение построено таким образом, чтобы помочь обучающимся заинтересоваться информатикой и найти ответы на вопросы, с которыми им приходится сталкиваться в повседневной жизни при работе с большим объемом информации; научиться общаться с компьютером, который ничего не умеет делать, если не умеет человек, углубить знания обучающихся в основах алгоритмизации и программирования.

Актуальность программы заключается в том, что она направлена на развитие у обучающихся алгоритмической культуры, дает базовые теоретические знания в области программирования, а также комплексы необходимых знаний в области информатизации и навыки, без которых не будет полноценного и осмысленного отношения обучающихся к техническому творчеству. Обучение направлено на развитие интереса к профессии программиста, что дает возможность студентам участвовать в конкурсах, фестивалях и других мероприятиях.

Новизна дополнительной общеобразовательной общеразвивающей программы - развитие навыков в области исследовательской и творческой технической деятельности.

Педагогическая целесообразность программы основывается на практическом освоении работы с технологиями информационного поиска и обработки больших данных, работы с инструментами анализа данных, основ математической статистики и теории вероятностей, основ математического моделирования, положительном опыте обучения алгоритмизации, развитии мышления, формировании научного мировоззрения.

Цель дополнительной общеобразовательной общеразвивающей программы: формирование у обучающегося аналитического мышления и, соответственно, знаний и умений, необходимых для успешного развития в отраслях, связанных со сложной аналитикой данных.

Задачи:

1. Обучающие:

- формировать у обучающихся навыки выполнения технологической цепочки разработки программ средствами языка программирования;
- углублять у обучающихся знания, умения и навыки решения задач по программированию и алгоритмизации;
- расширять и углублять знания в области информатики;
- приобщать детей к миру программирования;
- применять полученные знания в повседневной жизни;
- способствовать профессиональной ориентации обучающихся на выбор компьютерных специальностей.

2. Развивающие:

- развивать трудовые умения и навыки;
- развивать мелкую моторику;
- развивать творческую активность, любознательность, самостоятельность, целеустремленность,
- развивать мышление, память, внимание;
- развивать терпение, усидчивость.

3. Воспитательные:

- воспитывать трудолюбие, инициативу в работе;
- воспитывать умение работать в коллективе, помогать товарищу;
- воспитывать аккуратность при выполнении работ.

Отличительной особенностью данной программы является ее направленность на развитие мышления обучающихся и воспитания у них информационной культуры. На занятиях выполняются задания, развивающие творчество обучающихся, умение анализировать, систематизировать, визуализировать информацию. Обучающиеся учатся моделировать реально происходящие процессы, т.е. создавать информационную модель задачи.

Программа отличается от других программ тем, что:

- позволяет расширить информационную культуру;
- она ориентирована на развитие творческого потенциала и технических способностей обучающихся;
- практикоориентируемая и позволяет выполнить конкретную программу самостоятельно.

Тематическая направленность

В программу включен теоретический и практический материал, который наиболее интересен обучающимся объединения. Содержание практических работ и виды проектируемых объектов могут уточняться в зависимости от желания обучающихся, наличия материалов, средств и др.

С целью расширения кругозора обучающихся осуществлен переход от репродуктивного вида работ к самостоятельным, поисково-исследовательским видам деятельности. Поэтому основным методом обучения в данном элективном курсе является метод проектов, а основная методическая установка – обучение обучающиеся навыкам самостоятельной творческой деятельности.

Данная программа рассчитана на 1 год обучения.

Количество обучающихся в группах: 15 человек.

Режим занятий: 1 группа – 99 учебных часа, 3 (академических) часа в неделю, 2 группа- 96 часов, 3 академических часа в неделю, 3 группа -96 часов, 3 академических часа в неделю.

Этапы реализации программы

Первый этап – начальной подготовки:

- заложить основы алгоритмизации;
- научить обучающихся создавать проекты;
- обучить основам ООП

Второй этап – учебно-развивающий:

- сформировать основные навыки умения записи простых последовательностей действия;
- развить способности самостоятельного планирования пути достижения целей;
- научить оценивать правильность выполнения учебной задачи.
- подготовить конкретную работу.

Третий этап – усовершенствования:

- научить пользоваться языками программирования;
- познакомить с алгоритмами, которые используют в машинном обучении;

Форма и режим занятий

Основная форма занятий – *групповая или индивидуальная*, она связана со сменой деятельности. Продолжительность основного занятия составляет 2 часа с учетом здоровьесберегающих технологий: организационных моментов, упражнения, динамических пауз и т.д. Занятия обычно строятся по принципу: теоретический материал, основная часть, перерыв (гимнастика, паузы), повторение и закрепление.

Ожидаемые результаты

Будут знать	Будут уметь
Нереляционные базы данных	
<ul style="list-style-type: none"> • основные черты NoSQL баз данных; • типы хранилищ данных; • особенности документо-ориентированных СУБД; • особенности и возможности MongoDB; • язык запросов MongoDB 	<ul style="list-style-type: none"> • устанавливать и запускать MongoDB; • проводить моделирование данных в MongoDB; • выполнять запросы к базе MongoDB; • производить индексирование и оптимизацию запросов; • проводить репликации в MongoDB
Система распределенной обработки данных	
<ul style="list-style-type: none"> • Основные концепции Hadoop; • HDFS: Архитектуру, репликацию, чтение и запись данных; • Команды HDFS; • Структуру классической MapReduce-программы; 	<ul style="list-style-type: none"> • понимать ключевые концепции и архитектуру Hadoop; • уметь записывать и читать данные в/из HDFS; • разрабатывать программы для классического MapReduce и для YARN;

<ul style="list-style-type: none"> • Форматы данных для MapReduce; • Архитектуру YARN; • Выполнение работ в классическом MapReduce и в YARN; • Распределенный кеш и счетчики; • Hadoop Streaming; • Экосистему и вендоры Hadoop 	<ul style="list-style-type: none"> • работать с данными в HDFS из MapReduce-программ; • получить представление об экосистеме, сложившейся вокруг Hadoop® и ее ключевых компонентах
<p>Основы Data Science</p>	
<ul style="list-style-type: none"> • Свойства объектов Data Science • Содержание алгоритмов Data Science • регрессионного анализа, • кластерного анализа, • статистической классификации 	<ul style="list-style-type: none"> • Использовать высокоуровневые программные средства для реализации алгоритмов машинного обучения, строить математические модели, проверять статистическую адекватность моделей • Моделировать многомерные случайные данные, тестировать алгоритмы машинного обучения • Решать прикладные задачи Data Science с использованием методов регрессионного анализа, кластерного анализа и статистической классификации

Система отслеживания и оценивания результатов обучения обучающихся

Технические способности присутствуют в большей или меньшей степени у каждого здорового ребенка. В процессе обучения проводятся три вида диагностики: входящая, промежуточная и итоговая. Диагностика проводится в процессе работы и общения с обучающимися. Методы контроля и управления образовательным процессом - это наблюдение педагога в ходе занятий, анализ подготовки и участия воспитанников в различных мероприятиях колледжа, конкурсах.

Программа предусматривает реализацию двух ключевых компетентностей: социальную компетентность и компетентность в коммуникативной сфере.

Личностные результаты

- Формирование ответственного отношения к учению, готовности и способности, обучающихся к саморазвитию и самообразованию;
- развитие самостоятельности, личной ответственности за свои поступки;
- мотивация детей к познанию, творчеству, труду;
- формирование осознанного, уважительного и доброжелательного отношения к другому человеку;
- формирование коммуникативной компетентности в общении и сотрудничестве со сверстниками в процессе разных видов деятельности.

Метапредметные результаты

- Формирование умения самостоятельно определять цели своего обучения, ставить и формулировать для себя новые задачи в учёбе и познавательной деятельности;
- формирование умения самостоятельно планировать пути достижения целей, осознанно выбирать наиболее эффективные способы решения учебных и познавательных задач;
- формирование умения понимать причины успеха/неуспеха учебной деятельности;
- овладение различными способами поиска информации в соответствии с поставленными задачами;
- готовность слушать собеседника и вести диалог; излагать свое мнение и аргументировать свою точку зрения;
- формирование и развитие компетентности в области использования информационно-коммуникационных технологий.

Компетентностная модель

Будут сформированы следующие компетенции:

- **ценностно-смысловая компетенция** - (умение принимать решения, ставить цель и определять направление своих действий и поступков);
- **общекультурная компетенция** - (принимать и понимать точку зрения другого человека);
- **учебно-познавательная компетенция** - (самостоятельно находить материал, необходимый для работы, составлять план, оценивать и анализировать, делать выводы);
- **информационная компетенция** - (осваивать современные средства информации и информационные технологии);

- **коммуникативная компетенция** - (умение представлять себя и свою работу, отстаивать личную точку зрения, вести дискуссию, убеждать, задавать вопросы);
- выполнять работу над исследованием, учиться быть личностью, осознавать необходимость и значимость труда, который выполняешь - это и **социально-трудовая компетенция**, и **компетенция личностного самосовершенствования**.

2. Учебно - тематический план

№ п/п	Тема	Количество часов			Формы аттестации/контроля
		Всего	Теоретические занятия	Практические занятия	
Раздел 1. Нереляционные базы данных		33	5	28	
1.	Базы данных NoSQL	1	1		Входной контроль
2.	Введение в MongoDB. Устройство базы данных. Документы	1	1		
3.	Установка и администрирование базы данных	1		1	
4.	Добавление данных. Выборка из БД	1		1	
5.	Команды группировки.	1		1	
6.	Операторы выборки	1		1	
7.	Обновление и удаление данных	1		1	
8.	Установка ссылок в БД	1		1	
9.	Работа с индексами. Управление коллекцией	1		1	
10.	PHP и MongoDB	1	1		
11.	Установка драйвера для PHP. Подключение к MongoDB	1		1	
12.	Добавление данных. Выборка документов	1		1	
13.	Управление выборкой. Условные операторы	1		1	
14.	Обновление документов	1		1	
15.	Удаление данных	1		1	
16.	Работа с DBRef	1		1	
17.	Работа с GridFS	1		1	
18.	C# и MongoDB	1	1		
19.	Установка драйвера и подключение к базе данных.	1		1	
20.	Взаимодействие с коллекцией. Документы	1		1	

21	Модели данных. Настройка модели с помощью атрибутов	1		1	
22	Сохранение документов в базу данных	1		1	
23	Выборка из базы данных	1		1	
24	Фильтрация данных	1		1	
25	Интерфейс IFindFluent и его методью	1		1	
26	Агрегация и группировка	1		1	
27	Редактирование и удаление документов	1		1	
28	Метод BulkWriteAsync	1		1	
29	Работа с GridFS	1		1	
30	MongoDB и ASP.NET MVC. Создание проекта	1		1	
31	MongoDB и ASP.NET MVC. Работа с объектами	1		1	
32	MongoDB и ASP.NET MVC. Работа с файлами изображений	2		2	
Раздел 2. Система распределенной отработки данных		33	8	25	
1	Эволюция систем распределенных вычислений Hadoop.	1	1		
2	Концепция Data Lake	1		1	
3	Hadoop сервисы и основные компоненты: Name node	1	1		
4	Data Node. YARN	1		1	
5	Отказоустойчивость и высокая доступность	1		1	
6	Планировщик	1	1		
7	HDFS	1		1	
8	Установка 3х-узлового кластера	1		1	
9	Базовые операции с кластером Hadoop	1	1		
10	Блоки HDFS. Основные команды работы с HDFS.	1	1		
11	Операции чтения и записи, назначения HDFS. Архитектура HDFS	1		1	
12	Дисковые квоты	1	1		
13	Поддержка компрессии	1		1	
14	Базовые операции с файловыми операциями HDFS	1		1	
15	Основные форматы хранения данных TXT, AVRO, ORC, Parquet, Sequence файлы	1		1	
16	Импорт(загрузка) данных на HDFS	1		1	

17	Ведение в MapReduce.	1		1	
18	Компоненты MapReduce	1		1	
19	Работа программ MapReduce.	1		1	
20	YARN MapReduce v2	1		1	
21	Ограничения и параметры MapReduce и YARN	1	1		
22	Управление запуском пользовательских задач (jobs) под MapReduce	1		1	
23	Управление ресурсами и запуском задач с использованием YARN MapReduce	1		1	
24	Установка Hadoop кластера. Выбор начальной конфигурации	1	1		
25	Оптимизация уровня ядра для узлов. Начальная конфигурация HDFS и MapReduce	1		1	
26	Файлы логов и конфигураций. Установка Hadoop клиентов.	1		1	
27	Установка Hadoop кластера в облаке.	1		1	
28	Поиск узких мест. Производительность.	1		1	
29	Файловая система. Data Node. Сетевая производительность.	1		1	
30	FIFO scheduler	1		1	
31	Capacity scheduler (Планировщик по мощности)	1		1	
32	Fair scheduler (Гранулярное управление ресурсами)	2		2	
Раздел 3. Основы Data Science		33	8	25	
1.	Алгоритмы регрессии	1	1		
2.	Метод наименьших квадратов	1	1		
3.	Визуализация данных	1		1	
4.	Библиотека NumPy	1		1	
5.	Библиотека Matplotlib	1		1	
6.	Библиотека Seaborn	1		1	
7.	Библиотека Scikit-learn	1		1	
8.	Моделирование случайных данных	1		1	
9.	Регрессия в 1D на примерах модельных данных	1		1	
10.	Регрессия в 2D на примерах модельных данных	1		1	
11.	Статистический анализ значимости	1	1		

12	Оценка коэффициента корреляции	1		1	
13	Регрессия Ф. Гальтона	1		1	
14	Задача экстраполяции	1		1	
15	Построение многомерной регрессии	1	1		
16	Визуализация и корреляции данных в Boston House Prices Dataset	1		1	
17	Линейная аппроксимация данных в Boston House Prices Dataset	1		1	
18	Квадратичная аппроксимация данных в Boston House Prices Dataset	1		1	
19	Анализ эффектов взаимодействия предикторных переменных	1		1	
20	Регрессионные оценки в Boston House Prices Dataset	1		1	
21	Алгоритмы кластеризации	1	1		
22	Генераторы многомерных случайных данных	1		1	
23	Визуализация многомерных данных.	1		1	
24	Гистограммы данных	1	1		
25	Эмпирические кумулятивные распределения данных	1		1	
26	Прикладные задачи кластеризации	1		1	
27	Анализ Fisher's Iris Dataset	1		1	
28	Алгоритмы статистической классификации	1	1		
29	Реализации алгоритмов классификации в библиотеке Scikit-learn	1		1	
30	Прикладные задачи классификации. Анализ Dataset Titanic.	1		1	
31	Задача распознавания чисел. Проблема эффективности алгоритмов машинного обучения	1	1		
32	Компьютерный анализ эффективности алгоритмов машинного обучения. Проект	2		2	
Итого		99			

3. Содержание программы

Раздел 1. Нереляционные базы данных

Тема №1 Введение в нереляционные базы данных

Теория: Базы данных NoSQL

Тема №2 Работа с базой данных MongoDB

Теория: Введение в MongoDB. Устройство базы данных. Документы

Практика: Установка и администрирование базы данных. Добавление данных. Выборка из БД. Команды группировки. Операторы выборки. Обновление и удаление данных. Установка ссылок в БД. Работа с индексами. Управление коллекцией

Тема №3 PHP и MongoDB

Теория: PHP и MongoDB.

Практика: Установка драйвера для PHP. Подключение к MongoDB. Добавление данных. Выборка документов. Управление выборкой. Условные операторы. Обновление документов. Удаление данных. Работа с DBRef. Работа с GridFS

Тема №4 C# и MongoDB

Теория: C# и MongoDB

Практика: Установка драйвера и подключение к базе данных. Взаимодействие с коллекцией. Документы. Модели данных. Настройка модели с помощью атрибутов. Сохранение документов в базу данных. Выборка из базы данных. Фильтрация данных. Интерфейс IFindFluent и его методы. Агрегация и группировка. Редактирование и удаление документов. Метод BulkWriteAsync. Работа с GridFS. MongoDB и ASP.NET MVC. Создание проекта. MongoDB и ASP.NET MVC. Работа с объектами. MongoDB и ASP.NET MVC. Работа с файлами изображений

Раздел 2. Система распределенной отработки данных

Тема 2.1 Основы Hadoop и Big Data

Теория: Эволюция систем распределенных вычислений Hadoop, Hadoop сервисы и основные компоненты: Name node.

Практика: Концепция Data Lake, Data Node. YARN

Тема 2.2 Архитектура Apache Hadoop

Теория: Планировщик, Базовые операции с кластером Hadoop, Блоки HDFS. Основные команды работы с HDFS.

Практика: Отказоустойчивость и высокая доступность, HDFS, Установка 3х-узлового кластера, Операции чтения и записи, назначения HDFS. Архитектура HDFS.

Тема 2.3 Hadoop Distributed File System

Теория: Дисковые квоты

Практика: Поддержка компрессии, Базовые операции с файловыми операциями HDFS, Основные форматы хранения данных TXT, AVRO, ORC, Parquet, Sequence файлы, Импорт(загрузка) данных на HDFS.

Тема 2.4 MapReduce

Теория: Ведение в MapReduce, Ограничения и параметры MapReduce и YARN.

Практика: Компоненты MapReduce, Работа программ MapReduce, YARN MapReduce v2, Управление запуском пользовательских задач (jobs) под MapReduce, Управление ресурсами и запуском задач с использованием YARN MapReduce.

Тема 2.5 Архитектура YARN

Теория: Установка Hadoop кластера. Выбор начальной конфигурации

Практика: Оптимизация уровня ядра для узлов. Начальная конфигурация, HDFS и MapReduce, Файлы логов и конфигураций. Установка Hadoop клиентов, Установка Hadoop кластера в облаке, Поиск узких мест. Производительность., Файловая система. Data Node. Сетевая производительность., FIFO scheduler, Capacity scheduler (Планировщик по мощности), Fair scheduler (Гранулярное управление ресурсами)

Раздел 3. Основы Data Science

Тема 3.1 Регрессионный анализ

Теория: Алгоритмы регрессии. Метод наименьших квадратов. Статистический анализ значимости. Многомерная регрессия

Практика: Библиотеки NumPy, Matplotlib, Seaborn, Scikit-learn

Моделирование случайных данных. Регрессия в 1D и 2D на примерах модельных данных. Оценка коэффициентов корреляции. Регрессия Ф. Гальтона. Задача экстраполяции. Построение многомерной регрессии. Линейная и квадратичная аппроксимация данных и регрессионные оценки в Boston House Prices Dataset.

Тема 3.2 Кластерный анализ

Теория: Алгоритмы кластеризации

Практика: Генераторы многомерных случайных данных. Визуализация многомерных данных. Гистограммы данных. Эмпирические кумулятивные распределения данных. Реализации алгоритмов классификации в библиотеке Scikit-learn. Прикладные задачи кластеризации. Анализ Fisher's Iris Dataset

Тема 3.3 Статистическая классификация

Теория: Алгоритмы статистической классификации. Проблема эффективности алгоритмов машинного обучения

Практика: Реализации алгоритмов классификации в библиотеке Scikit-learn. Прикладные задачи классификации. Анализ Dataset Titanic. Задача распознавания чисел. Компьютерный анализ эффективности алгоритмов машинного обучения

Способы диагностики и контроля результатов

Диагностика: первичная (на первом занятии), итоговая (в конце каждого раздела).

Формы подведения итогов реализации образовательной программы

1. Итоговые занятия.
2. Зачетные занятия и сдача работ.
5. Отчетные в конце каждого раздела.

4. Методическое обеспечение программы

В работе предполагается использование разнообразных методов и приемов как традиционных (словесные, наглядные, практические), так и новаторских (разнообразные "модели" занятий, применение здоровьесберегающих технологий, мониторинг и диагностика способностей, цифровых компьютерных технологий, система нетрадиционных творческих заданий).

Формы занятий:

- общие занятия (ведется теоретическая работа, закрепление пройденного материала, воспитательные беседы);
- групповые (ведется практическая работа).

Структура занятия:

- организационное начало - установка на занятие, постановка задач;
- повторение предыдущей темы;
- ознакомление и работа с новым заданием;
- заключение, итог.

5. Материально-техническое обеспечение

1. Нереляционные базы данных

- 1.1. Лаборатория
- 1.2. Персональные компьютеры по количеству обучающихся
- 1.3. Посадочные места по количеству обучающихся
- 1.4. Программное обеспечение: MongoDB, Visual Studio

6. Система распределенной обработки данных

- 2.1 Лаборатория
- 2.2 Персональные компьютеры
- 2.3 Посадочные места по числу обучающихся
- 2.4 Программное обеспечение: Virtual Box 6.0, Linux Ubuntu, Apache Hadoop

7. Основы Data Science

- 7.1. Лаборатория
- 7.2. Персональные компьютеры
- 7.3. Посадочные места по числу обучающихся
- 7.4. Программное обеспечение Anaconda, WinPython, PyCharm

Список литературы

1. Нереляционные базы данных

1.1 MongoDB в действии. / Пер. с англ. Слинкина А.А. – М.: ДМК Пресс, 2017. – 394 с.: ил.

2. Система распределенной обработки данных

2.1. Hadoop в действии, Чак Лэм - ДМК Пресс, 2017.

2.2 Hadoop. Подробное руководство, Том Уайт - Питер, 2016.

2.3 Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных, Петер Флах - ДМК Пресс, 2015.

3. Основы Data Science

3.1. Грас Дж. Data Science. Наука о данных с нуля: Пер. с англ. – СПб.: БХВ-Петербург, 2017. 336 с.: ил.

3.2. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. Пер. с англ.: – СПб.: ООО «Альфа-книга», 2018. – 480 с.: ил

Интернет-ресурсы

Нереляционные базы данных

<https://metanit.com/nosql/mongodb/>

Система распределенной обработки данных

1. Официальный сайт Hadoop - <https://hadoop.apache.org/>

2. Apache Hadoop в Amazon EMR - <https://aws.amazon.com/ru/emr/features/hadoop/>
3. Подробнее о Hadoop - <https://habr.com/post/240405/>
4. История Hadoop - <https://ru.wikipedia.org/wiki/Hadoop>
5. Основы Data Science
<https://scikit-learn.org/stable/index.html>

Кадровое обеспечение программы

Программа реализуется педагогами дополнительного образования, имеющим профессиональное образование в области, соответствующей профилю программы, и постоянно повышающим уровень профессионального мастерства.

