

Образовательное частное учреждение  
дополнительного профессионального образования  
«Центр компьютерного обучения «Специалист.Ру»  
Учебно-научного центра при МГТУ им. Н.Э. Баумана»  
(ОЧУ «Специалист.Ру»)

123317, город Москва, улица Зоологическая, дом 11, строение 2, помещение 1, этаж 2, комната 14  
ИНН 7701345493, ОГРН 1037701927031

Утверждаю:

Директор ОЧУ «Специалист.Ру»



\_\_\_\_\_/О.В.Пичугина/

10 \_\_\_\_ 2019\_\_ года

Дополнительная профессиональная программа  
повышения квалификации  
«Основы работы с большими данными  
(Data Science)»

город Москва

Программа разработана в соответствии с приказом Министерства образования и науки Российской Федерации от 1 июля 2013 г. N 499 "Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам" и Федеральным законом от 29.12.2012 N 273-ФЗ (ред. от 07.03.2018) "Об образовании в Российской Федерации".

Повышение квалификации слушателей, осуществляемое в соответствии с программой, проводится с использованием модульного принципа построения учебного плана с применением различных образовательных технологий, в том числе дистанционных образовательных технологий и электронного обучения в соответствии с законодательством об образовании.

Дополнительная профессиональная программа повышения квалификации, разработана образовательной организацией в соответствии с законодательством Российской Федерации, включает все модули, указанные в учебном плане.

Содержание оценочных и методических материалов определяется образовательной организацией самостоятельно с учетом положений законодательства об образовании Российской Федерации.

Структура дополнительной профессиональной программы соответствует требованиям Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденного приказом Минобрнауки России от 1 июля 2013 г. N 499.

Объем дополнительной профессиональной программы вне зависимости от применяемых образовательных технологий, должен быть не менее 16 академических часов. Сроки ее освоения определяются образовательной организацией самостоятельно.

Формы обучения слушателей (очная, очно-заочная, заочная) определяются образовательной организацией самостоятельно.

К освоению дополнительных профессиональных программ допускаются:

- лица, имеющие среднее профессиональное и (или) высшее образование;
- лица, получающие среднее профессиональное и (или) высшее образование.

Для определения структуры дополнительной профессиональной программы и трудоемкости ее освоения может применяться система зачетных единиц. Количество зачетных единиц по дополнительной профессиональной программе устанавливается организацией.

Образовательная деятельность слушателей предусматривает следующие виды учебных занятий и учебных работ: лекции, практические и семинарские занятия, лабораторные работы, круглые столы, мастер-классы, мастерские, деловые игры, ролевые игры, тренинги, семинары по обмену опытом, выездные занятия, консультации, выполнение аттестационной, дипломной, проектной работы и другие виды учебных занятий и учебных работ, определенные учебным планом.

### 1. Цель программы:

В результате прохождения обучения слушатель получит знания и навыки знакомства с современными подходами к анализу «больших данных»..

### Совершенствуемые компетенции

№	Компетенция	Направление подготовки		
		ФГОС	ВО	ПО
		НАПРАВЛЕНИЮ ПОДГОТОВКИ 090302 «ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ»		

		(УРОВЕНЬ БАКАЛАВРИАТА"
		Код компетенции
1	способностью проводить предпроектное обследование объекта проектирования, системный анализ предметной области, их взаимосвязей	ПК-1
2	способностью проводить техническое проектирование	ПК-2
3	способностью проводить рабочее проектирование	ПК-3
4	способностью проводить выбор исходных данных для проектирования	ПК-4
5	способностью оценивать надежность и качество функционирования объекта проектирования	ПК-6
6	способность к проектированию базовых и прикладных информационных технологий	ПК-11
7	способностью проводить сбор, анализ научно-технической информации, отечественного и зарубежного опыта по тематике исследования	ПК-22
8	готовностью участвовать в постановке и проведении экспериментальных исследований	ПК-23
9	способностью обосновывать правильность выбранной модели, сопоставляя результаты экспериментальных данных и полученных решений	ПК-25
10	способностью использовать математические методы обработки, анализа и синтеза результатов профессиональных исследований	ПК-26
11	способностью формировать новые конкурентоспособные идеи и реализовывать их в проектах	ПК-27
12	способностью поддерживать работоспособность информационных систем и технологий в заданных функциональных характеристиках и соответствии критериям качества	ПК-30
13	способностью составлять инструкции по эксплуатации информационных систем	ПК-33

**Совершенствуемые компетенции в соответствии с трудовыми функциями профессионального стандарта «Менеджер по информационным технологиям», утвержденного приказом Минтруда и социальной защиты РФ от 13 октября 2014 г. N 716н**

№	Компетенция	Направление подготовки
		ПРОФЕССИОНАЛЬНЫЙ СТАНДАРТ " «Менеджер по информационным технологиям», утвержденного приказом Минтруда и социальной защиты РФ от 13 октября 2014 г. N 716н
		Трудовые функции (код)
14	<b>Управление ИТ-инновациями</b>	D/01.9 -Управление формированием вклада ИТ в создание и реализацию инновационной стратегии D/03.9 -Управление оценкой эффективности ИТ-инноваций

D/04.9 - Управление знаниями с помощью ИТ
---

D/05.9 - Управление взаимоотношениями с заинтересованными лицами
--

### **Планируемый результат обучения:**

#### **После окончания обучения Слушатель будет знать:**

понимать концепцию больших данных (Big Data) ;  
знать типовые задачи, для решения которых применяется машинное обучение: анализ трендов, социальных сетей; распознавание графических, видео- и аудио-образов, текста; прогностика действий (на примере покупок)  
владеть базовыми математическими понятиями  
понимать основные методы обработки и анализа данных: регрессия, нейронные сети; графы, К-значные логики  
знать основные современные инструменты анализа данных  
понимать принципы организации и структуру команд по работе с бигдата

#### **После окончания обучения Слушатель будет уметь:**

- определять источники сбора информации и формировать требования к ним;
- Применять стандартный CRISP-DM процесс для Вашей организации
- подбирать команду для работы с большими данными (Big Data);
- выбирать инструментарий для практической работы;
- Применять специализированные инструменты Excel – «Пакет анализа данных» и «Тренды»
- Применять «дерево решений»
- Определять подходящие инструменты и методы для решения основных классов задач машинного обучения и взаимодействовать с разработчиками
- Использовать методы классификации данных для машинного обучения
- Подбирать выборки разработки, тестовую и обучающую для достижения наилучших результатов анализа информации
- Организовывать реорганизацию работы компании для применения управления на основе больших данных

Данный курс соответствует требованиям профессионального стандарта **«Менеджер по информационным технологиям»**, утвержденного приказом Минтруда и социальной защиты РФ от 13 октября 2014 г. N 716н

#### **Учебный план:**

Категория слушателей: курс предназначен для:

- Руководители компаний и подразделений
- Линейные менеджеры
- Бизнес-аналитики
- Разработчики

Другие сотрудники, вовлеченные в аналитическую деятельность компании  
Требования к предварительной подготовке:

Успешное окончание курса: Эксель1

Срок обучения: 16 академических часов, 8 самостоятельно

Форма обучения: очная, очно-заочная, заочная. По желанию слушателя форма обучения может быть изменена и/или дополнена.

Режим занятий: дневной, вечерний, группы выходного дня.

№ п/п	Наименование модулей по программе	Общая трудоемкость (акад. часов)	Всего ауд. ч	В том числе		СРС, ч	ПА*	
				Лекций	Практических занятий			
1	<b>Модуль 1. Область применения больших данных. Постановка задачи.</b>	2	1	1	0	1	Уст. пров	
2	<b>Модуль 2. Data Science: назначение, специальности</b>	3	2	1	1	1	Уст. пров	
3	<b>Модуль 3. Сбор и подготовка исходных данных. Визуализация</b>	3	2	1	1	1	Уст. пров	
4	<b>Модуль 4. Основы математической статистики. ANOVA</b>	4	3	1	2	1	Уст. пров	
5	<b>Модуль 5. Регрессионный анализ. Прогностика и тренды</b>	4	3	1	2	1	Уст. пров	
6	<b>Модуль 6. Задача классификации данных. Нейронные сети. Machine Learning</b>	4	3	1	2	1	Уст. пров	
7	<b>Модуль 7. Другие инструменты анализа данных</b>	2	1	0	1	1	Уст. пров	
8	<b>Модуль 8. Выводы и рекомендации по специализации</b>	2	1	0	1	1		
	Итого:	24	16	6	10	8		
	Итоговая аттестация	тестирование						

Для всех видов аудиторных занятий академический час устанавливается продолжительностью 45 минут. Форма промежуточной аттестации – см. в ЛНА «Положение о проведении промежуточной аттестации слушателей и осуществлении текущего контроля их успеваемости» п.3.3

## 2. Календарный учебный график

Календарный учебный график формируется при осуществлении обучения в течение всего календарного года. По мере набора групп слушателей по программе составляется календарный график, учитывающий объемы лекций, практики, самоподготовки, выезды на объекты.

Неделя обучения	1	2	3	4	5	6	7	Итого часов
	пн	вт	ср	чт	пт	сб	вс	
1 неделя	8	8АИ	-	-	-	-	-	16
СРС	4	4						8
Итого:								24

Примечание: ИА – Итоговая аттестация (тестирование)

### 3. Рабочие программы учебных предметов

#### Модуль 1. Область применения больших данных. Типовые задачи.

- Цели курса
- Определение основных понятий
- История науки о данных
- Выгоды от работы с большими данными
- Типовые задачи: прогноз продаж, производства, спроса. Анализ поведения. Распознавание образов. Экспертные системы.

#### Модуль 2. Сбор и подготовка исходных данных. Методика CRISP-DM

- С чего начать. Межотраслевая стандартная методика работы с данными CRISP-DM
- Описательное и ассоциативное исследование исходных данных
- Сегментирование и очистка данных (slice and dice). Примеры инструментов Excel
- Визуализация данных в Excel. Как использовать сводные таблицы и диаграммы.
- Практическая работа. Сегментировать и очистить тестовый набор данных

#### Модуль 3. Основы математической статистики, ANOVA. Настройка Excel «Пакет анализа»

- Описательная статистика
- Среднее, наиболее вероятное, медиана
- Дисперсия, стандартное отклонение, стандартная ошибка
- Виды распределений
- Пакет анализа данных Excel
- Обзор других прикладных средств работы с данными (R, Python, Octave, MathLab, специализированные БД)
- Практическая работа. Определить статистические характеристики выборки данных.

#### Модуль 4. Задача прогноза продаж. Понятие машинного обучения. Корреляция. Регрессионный анализ

- Постановка задачи оценки взаимосвязи между различными факторами и построение прогноза
- Корреляция. Коэффициент Пирсона
- Критерий Стьюдента (Т-анализ)
- Основы машинного обучения
- Регрессионный анализ
- Критерий Фишера
- Построение и анализ трендов в Excel
- Практическая работа. Определить наличие корреляции и регрессионную зависимость между двумя выборками данных. Построить тренд.

#### Модуль 5. Задачи классификации и распознавания образов, видео, речи, текста. Понятие нейронных сетей. Примеры применения

- Задача сегментации дискретных данных на примере задач распознавания (графика, речь, текст)
- Нейронные сети как инструмент решения задач классификации
- Демонстрация на примерах Azure, AWS

**Модуль 6. Задача исследования социальных сетей. Задача прогнозирования поведения пользователя. Социальные и направленные графы. Деревья решений. Примеры применения**

- Задачи классификации данных в социальных сетях и поиска оптимального решения (маршрута)
- Графы как инструмент решения задач на социальных графах и прогнозирования поведения
- Дерево решений
- Разбиение на выборки (обучающую, тестовую, проверочную)
- Анализ ошибок обучения. Базис и отклонения. Ручная корректировка
- Практическая работа: провести классификацию набора данных и его разбиение на сегменты

**Модуль 7. Продвинутое инструменты: глубокое машинное обучение, искусственный интеллект, нечеткие множества**

- Понятие Deep Machine Learning
- Многофакторный бизнес анализ на примере нечетких логик

**Модуль 8. Профориентация по специальностям в Data Science. Выводы и рекомендации по построению и организации работы команды**

- Роли специалистов по DS: аналитик данных, ученый по данным, программист, цифровой директор
- Требования к компетенциям и взаимодействию сотрудников в области аналитики данных
- Состав и требования к проектной команде для DS
- Подготовка компании к применению «бигдаты»

**Организационно-педагогические условия**

Соблюдение требований к кадровым условиям реализации дополнительной профессиональной программы:

а) преподавательский состав образовательной организации, обеспечивающий образовательный процесс, обладает высшим образованием и стажем преподавания по изучаемой тематике не менее 1 года и (или) практической работы в областях знаний, предусмотренных модулями программы, не менее 3 (трех) лет;

б) образовательной организацией наряду с традиционными лекционно-семинарскими занятиями применяются современные эффективные методики преподавания с применением интерактивных форм обучения, аудиовизуальных средств, информационно-телекоммуникационных ресурсов и наглядных учебных пособий.

Соблюдение требований к материально-техническому и учебно-методическому обеспечению дополнительной профессиональной программы:

а) образовательная организация располагает необходимой материально-технической базой, включая современные аудитории, библиотеку, аудиовизуальные средства обучения, мультимедийную аппаратуру, оргтехнику, копировальные аппараты. Материальная база соответствует санитарным и техническим нормам и правилам и обеспечивает проведение всех видов практической и дисциплинарной подготовки слушателей, предусмотренных учебным планом реализуемой дополнительной профессиональной программы.

б) в случае применения электронного обучения, дистанционных образовательных технологий каждый обучающийся в течение всего периода обучения обеспечивается

индивидуальным неограниченным доступом к электронной информационно-образовательной среде, содержащей все электронные образовательные ресурсы, перечисленные в модулях дополнительной профессиональной программы.

#### **4. Формы аттестации и оценочные материалы**

Образовательная организация несет ответственность за качество подготовки слушателей и реализацию дополнительной профессиональной программы в полном объеме в соответствии с учебным планом.

Оценка качества освоения дополнительной профессиональной программы слушателей включает текущий контроль успеваемости и итоговую аттестацию.

Результаты итоговой аттестации слушателей ДПП в соответствии с формой итоговой аттестации, установленной учебным планом, выставляются по двух бальной шкале («зачтено\незачтено»).

Слушателям, успешно освоившим дополнительную профессиональную программу и прошедшим итоговую аттестацию, выдается удостоверение о повышении квалификации.

Слушателям, не прошедшим итоговой аттестации или получившим на итоговой аттестации неудовлетворительные результаты, а также лицам, освоившим часть дополнительной профессиональной программы и (или) отчисленным из образовательной организации, выдается справка об обучении или о периоде обучения по образцу, самостоятельно устанавливаемому образовательной организацией.

Итоговая аттестация проводится по форме тестирования в соответствии с учебным планом. Результаты итоговой аттестации заносятся в соответствующие документы

#### **Вопросы промежуточной аттестации**

- Где применяют большие данные?
- Как использовать сводные таблицы и диаграммы?
- Какие бывают виды распределений?
- Назовите Коэффициент Пирсона?
- Нейронные сети как инструмент решения задач классификации?
- Как происходит разбиение на выборки?

## **Условия прохождения**

**Время(мин):** 30

**Количество вопросов:** 12

**Проходной балл(ПБ):** 7

**ПБ средний уровень:** 9

**ПБ эксперт:** 10

### **Модули**

**Область применения больших данных. Постановка задачи** - 5%

**Data Science: назначение, области применения** - 5%

**Сбор и подготовка исходных данных** - 20%

**Основы математической статистики** - 25%

**Визуализация данных** - 5%

**Инструменты анализа данных** - 35%

**Выгоды для компании от применения аналитики больших данных (Big Data) - 5%**

## **Data Science: назначение, области применения**

### **Вопрос 1/20**

Роль аналитика данных предполагает в первую очередь знание какой области?

**Выберите один ответ:**

- Программирование
- Предметная область
- Алгоритмы машинного обучения
- Математическая статистика

### **Вопрос 2/20**

Ученый по данным должен в большей степени владеть навыками в областях (выберите все подходящие)

**Выберите несколько ответов:**

- Регрессионный анализ
- Предметная область
- Алгоритмы машинного обучения
- Программирование

## **Основы математической статистики**

### **Вопрос 3/20**

Найдите среднее значение ряда: 12, 12, 14, 18, 25, 38, 42, 79

**Выберите один ответ:**

- 22
- 30
- 8
- 12

### **Вопрос 4/20**

Найдите медиану ряда: 12, 12, 14, 18, 25, 38, 42, 79

**Выберите один ответ:**

- 30
- 22

- 12
- 25

### Вопрос 5/20

Найдите моду ряда: 12, 12, 14, 18, 25, 38, 42, 79

Выберите один ответ:

- 30
- 22
- 25
- 12

### Вопрос 6/20

Выясните, есть ли корреляция между рядами А и Б. Ряд А: 25 30 35 32 30 30 25 28 32 42 39 35 26 32 38 31 29 31 21 30 35 47 48 36 Ряд Б: 4.8 4.70 4.90 4.70 4.60 4.70 4.80 4.80 4.90 4.80 4.60 4.50 4.80 4.80 4.80 4.90 4.70 4.75 4.85 4.80 4.80 4.75 4.60 4.50

Выберите один ответ:

- Корреляция есть, т.к.  $R$  положительный
- Корреляции нет, т.к.  $R$  отрицательный
- Корреляции нет, т.к.  $R < 0.5$
- Корреляции нет, т.к.  $R < -0.5$

### Вопрос 7/20

Ряды А и Б содержат по 12 элементов каждый. Каково число степеней свободы при линейной регрессии между этими рядами?

Выберите один ответ:

- 1
- 22
- 23
- 24

## Область применения больших данных. Постановка задачи

### Вопрос 8/20

Что из перечисленного служит математической мерой корреляции случайных величин?

Выберите один ответ:

- Линейная регрессия

- Медиана
- Коэффициент Фишера
- Мода

### Вопрос 9/20

"Совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети" называется \_\_\_\_\_

**Выберите один ответ:**

- Большие данные
- Математическая регрессия
- Линейная регрессия
- Искусственный интеллект

### Вопрос 10/20

"математическое выражение, отражающее зависимость зависимой переменной  $y$  от независимых переменных  $x$  при условии, что это выражение будет иметь статистическую значимость" называется \_\_\_\_\_

**Выберите один ответ:**

- Регрессией
- Линейной регрессией
- Машинным обучением
- Искусственным интеллектом

## Сбор и подготовка исходных данных

### Вопрос 11/20

Рассчитайте стандартную ошибку для ряда: 25 30 35 32 30 30 25 28 32 42 39 35 26 32 38 31 29 31 21 30 35 47 48 36

**Выберите один ответ:**

- 32.79
- 1.34
- 6.57
- 43.12

### Вопрос 12/20

Рассчитайте стандартное отклонение для ряда: 25 30 35 32 30 30 25 28 32 42 39 35 26 32 38 31 29 31 21 30 35 47 48 36

**Выберите один ответ:**

- 1.34
- 43.12
- 32.79
- 6.57

### **Вопрос 13/20**

Вы анализируете распределение продаж по месяцам. Какой тип анализа, наиболее вероятно, Вы применяете?

**Выберите один ответ:**

- Статистический
- Сравнительный
- Описательный
- Ассоциативный

### **Вопрос 14/20**

Вы анализируете динамику продаж ежемесячно в сравнении с тремя предшествующими годами. Вы выполняете \_\_\_\_\_ анализ

**Выберите один ответ:**

- Регрессионный
- Описательный
- Сравнительный
- Статистический

## **Инструменты анализа данных**

### **Вопрос 15/20**

Рассчитайте коэффициент Фишера для рядов А и Б: Ряд А: 25 30 35 32 30 30 25 28 32 42 39 35 26 32 38 31 29 31 21 30 35 47 48 36 Ряд Б: 4.8 4.70 4.90 4.70 4.60 4.70 4.80 4.80 4.90 4.80 4.60 4.50 4.80 4.80 4.80 4.90 4.70 4.75 4.85 4.80 4.80 4.75 4.60 4.50

**Выберите один ответ:**

- 0.32
- 0.12
- 2.55

- 0.1

### Вопрос 16/20

Рассчитайте коэффициент Стьюдента при двустороннем распределении для рядов А и Б и вероятности нулевой гипотезы 5%: А: 25 30 35 32 30 30 25 28 32 42 39 35 26 32 38 31 29 31 21 30 35 47 48 36 Б: 4.8 4.70 4.90 4.70 4.60 4.70 4.80 4.80 4.90 4.80 4.60 4.50 4.80 4.80 4.80 4.90 4.70 4.75 4.85 4.80 4.80 4.75 4.60 4.50

**Выберите один ответ:**

- 3.87E-25
- 1.93E-25
- 1.68
- 2.01

### Вопрос 17/20

Оцените, есть ли корреляция между рядами А и Б и вероятности нулевой гипотезы 5% и объясните, почему: А: 25 30 35 32 30 30 25 28 32 42 39 35 26 32 38 31 29 31 21 30 35 47 48 36 Б: 4.8 4.70 4.90 4.70 4.60 4.70 4.80 4.80 4.90 4.80 4.60 4.50 4.80 4.80 4.80 4.90 4.70 4.75 4.85 4.80 4.80 4.75 4.60 4.50

**Выберите один ответ:**

- Неизвестно, т.к. количество наблюдений слишком мало
- Есть, т.к. коэф-т корреляции  $R=0.78$
- Есть, т.к. коэффициент Стьюдента для двустороннего распределения равен 3.87E-25
- Нет, т.к. коэф-т корреляции  $R$  меньше нуля

### Вопрос 18/20

Какие виды обучения алгоритмов Вы знаете? Выберите все подходящие варианты.

**Выберите несколько ответов:**

- С супервизором
- Без учителя
- Искусственный интеллект
- Машинное обучение

## Визуализация данных

### Вопрос 19/20

Инструмент визуализации данных в Excel называется

**Выберите один ответ:**

- Pivot
- ANOVA
- R-анализ
- Пакет анализа

## **Выгоды для компании от применения аналитики больших данных (Big Data)**

### **Вопрос 20/20**

Как называется модель внедрения новых подходов в организации?

#### **Выберите один ответ:**

- Восьми шагов
- Бонус-малус
- Коттера
- Визионерство