

## Программа учебной дисциплины «Введение в Data Science»

Утверждена

Академическим советом ООП

Протокол № 3 от «29» августа 2018 г.

<b>Автор</b>	Артюшина Е.В., Привалов И.В.
<b>Число кредитов</b>	3
<b>Контактная работа (час.)</b>	40
<b>Самостоятельная работа (час.)</b>	74
<b>Курс</b>	1 курс
<b>Формат изучения дисциплины</b>	Без использования онлайн курса

### I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Введение в Data Science» являются: формирование знаний, умений и навыков, а также развитие компетенций, необходимых для формирования у будущих менеджеров data-analyticthinking, с помощью которого они смогут получать знания и инсайт из данных. Фундаментальные принципы DataScience рассматриваются с точки зрения решения бизнес проблем.

В результате освоения дисциплины студент должен

**знать:**

- основы машинного обучения и типовые задачи,
- типовые информационные системы, применяемые в современном бизнесе,
- основы аналитики данных от информационных систем,
- основные концепты статистики.

**уметь:**

- работать с базами данных, создавать и перекодировать переменные,
- уметь применять основные методы и процедуры трансформации данных;
- использовать информационные системы при принятии управленческих решений;

**владеть:**

- адекватными методами обработки данных в зависимости от их структуры и решаемой управленческой задачи,
- навыками работы с программным обеспечением для обработки и хранения данных.

Изучение дисциплины «Введение в Data Science» базируется на следующих дисциплинах:

- Информатики в рамках школьного курса.;
- Математика.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Знаниями из области информатики школьного курса.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин, ориентированных на освоение аналитических, IT-компетенций в управленческой деятельности:

- Качественные и количественные методы разработки и принятия управленческих решений,
- Методы научных исследований в менеджменте
- Математическое моделирование бизнес-процессов
- Базы данных.

## **II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ**

### **Раздел 1. Введение.**

Что такое DataScience. Примеры проектов из практической деятельности. Описание ролей и задействованных специалистов. Этапы проекта. Разбор примера плана проекта с указанием ролей, задач, проблем. Рынок данных. Юридические аспекты работы с данными. Инструменты аналитика — python, R, ноутбуки; SAS, SPSS, RapidMiner; библиотеки — какие и зачем; Hadoop, Spark.

### **Раздел 2. Современные информационные технологии в бизнесе. Источники данных в бизнесе.**

Информационные системы в менеджменте как источники данных. Типы данных в информационных системах. DataMining и ProcessMining

### **Раздел 3. Данные, базы данных, большие данные.**

Данные как ключевой компонент в MIS. Базы данных, структура базы данных и системы управления базами данных (СУБД). Моделирование данных. Реляционные и иерархические базы данных. Нормализация. SQL. Выбор СУБД. Большие данные.

### **Раздел 4. Поисковые технологии и информационная безопасность.**

Понятие информационной безопасности. Законодательство в области ИБ. Индексация. Оценка качества работы информационно-поисковых систем. Обработки запросов на поиск в индексе. Модель векторного пространства. Ранжирование.

### **Раздел 5. Возможности пакета EXCEL в обработке экономических данных.**

Меню и панели инструментов. Основные типы данных. Абсолютная и относительная адресация. Копирование. Работа с рядами. Форматы данных. Числовое форматирование. Создание пользовательских форматов. Форматирование таблицы. Функции и формулы в пакете Excel. Классы табличных функций: математические, статистические, логические, финансовые, даты и времени и др. Ввод функций и формул. Мастер функций. Редактирование формул. Основы визуализации данных. Типы диаграмм. График, точечная, гистограммы. Редактирование диаграмм Форматирование диаграмм.

### **Раздел 6. Пакет анализа и прикладное программирование в среде EXCEL**

Поиск решения средствами EXCEL. Решение задачи подбора параметра в ряде экономических задач. Макросы в пакете EXCEL. Язык VisualBasic. Назначение и использование макросов. Создание макроса в виде последовательности

команд. Логические выражения в макросах. Автоматизация приложений с помощью макросов. Переменные и константы. Функции и подпрограммы. Управление выполнением программы.

### III. ОЦЕНИВАНИЕ

*Текущий контроль:*

*Работа студентов на лекционных и семинарских занятиях.* Преподаватель оценивает работу студентов на семинарских занятиях: активность на семинарских занятиях в дискуссиях, правильность решения задач; по активности студентов в дискуссиях, в ответах на вопросы, в разборах кейсов, решении задач, групповой работе.. Результаты складываются в оценку за аудиторную работу. Оценки за работу на семинарских занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на семинарских занятиях определяется перед итоговым контролем - *О<sub>аудиторная</sub>*.

*Итоговый контроль* – экзамен. Форма экзамена: Письменное тестирование, практические задачи 120 мин.. По итогам проведения тестирования для получения оценки "10 баллов" студенту необходимо правильно выполнить 90-100% заданий. Для получения оценки "8-9 баллов" – не менее 80%. "7 баллов" – 70%. "5-6 баллов" – не менее 50% поставленных задач. Менее 50% правильных ответов – оценка ниже 5-ти баллов

#### **Порядок формирования оценок по дисциплине**

Результирующая оценка за дисциплину рассчитывается следующим образом,  $O_{\text{экс}}$  – оценка за контрольную работу непосредственно на экзамене:

$$O_{\text{результ}} = 0,4 * O_{\text{экс}} + 0,6 * O_{\text{аудиторная}}$$

Способ округления накопленной оценки итогового контроля в форме экзамена: арифметический.

В диплом выставляется результирующая оценка по учебной дисциплине.

### IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

#### **Оценочные средства для текущего контроля студента**

*(образец, может корректироваться под задачу и ситуацию)*

1. Нужно посчитать стоимость за сутки проживания в отеле в Еврозоне. Все цены представлены в валюте евро конвертируются в рубли по курсу 1 EUR = 74 RUB. Курсы валют изменяются каждый день, поэтому нужно сделать прайс так, чтобы можно было легко вычислить актуальные цены в рублях. Цены вычисляются с помощью формул с абсолютными или смешанными адресами ячеек так, чтобы не нужно было изменять в них формулы при каждом изменении курса валют. Изменения будут вноситься только в одну ячейку.

2. Допустим, Вы хотите приобрести автомобиль в кредит. Максимальная сумма ежемесячного взноса, которую Вы можете себе позволить, составляет 700\$. Банк не может выдать Вам кредит сроком более чем на 3 года, с процентной ставкой 5,5% годовых. Можете ли вы себе позволить при таких условиях кредитования приобрести автомобиль стоимостью в 30 000\$, а если нет, то на какую сумму

можно рассчитывать?

### Вопросы для освоения качества дисциплины.

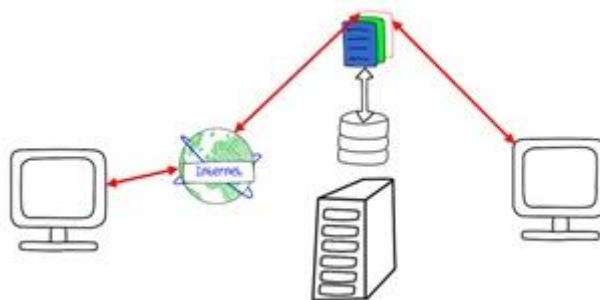
*Итоговый тест (образец, может корректироваться под задачу и ситуацию)*

### Понятие информационных систем в менеджменте, варианты внедрения информационных систем для менеджмента на предприятие.

1. Внедренные готовых (пакетных) решений МИС имеют конкурентные преимущества.
  - Правда
  - Ложь
2. Объясните преимущества заказных (разработанных на заказ) информационных систем над готовыми пакетными программными решениями:
  - Они дешевле.
  - Они включают лучшие практики индустрии.
  - Их обслуживание легче.
  - Они лучше соответствуют конкретным бизнес-требованиям.
3. Обоснуйте выбор информационной системы, если вы хотите открыть микро бизнес по пошиву одежды на дому и вам необходима информационная система для бухгалтерского учёта

Распределенные информационные системы, инфраструктуры и архитектуры

1. Локальные сети не могут включать:
  - Такие устройства, как принтеры.
  - Серверы.
  - ПК, подключенные через карты Ethernet.
  - Устройства, расположенные очень далеко (например, в разных странах) друг от друга
2. Объясните структуру распределенные информационных систем:
  - Есть центральный компьютер (ы) и терминалы / клиентские компьютеры; первые имеют программное обеспечение, а последние - данные.
  - Есть центральный компьютер (ы) и терминалы / клиентские компьютеры; у первых есть программное обеспечение, и у обоих есть данные.
  - Есть центральный компьютер (ы) и терминалы / клиентские компьютеры; оба могут иметь программное обеспечение и данные. правильный
  - Есть центральный компьютер (ы) и терминалы / клиентские компьютеры; только первые имеют как программное обеспечение, так и данные; последние не имеют ни того, ни другого.



3. Определите тип архитектуры

### Данные в информационных системах базы данных, большие данные

1. Закон Мура также может применяться к Большим Данным.
  - Правда
  - Ложь

2. В базе данных создается таблица «people», и там будут храниться следующие данные: имя, фамилия, высота, вес, персональный идентификатор, возраст, пол. Какое поле будет являться лучшим Первичным ключом?
  - Фамилия
  - Имя
  - Персональный идентификационный номер
  - Возраст
3. Для чего используется SQL?

### **Интегрированные информационные системы (ERP) и интегрированные бизнес-процессы**

1. ERP - это интегрированные информационные системы (они объединяют модули, специализированные в различных бизнес-функциях). Однако, поскольку каждая бизнес-функция (например, Accounting, Production ...) имеет разные потребности, каждый модуль использует свои собственные данные и базу данных.
  - Правда
  - Ложь
2. В настоящее время компания использует систему ERP, охватывающую три модуля (учет, производство и продажи), установленные в прошлом году. Компания решила расширить свою реализацию ERP, чтобы дополнительно интегрировать следующие бизнес-функции: управление запасами и человеческие ресурсы. Выберите вариант развития событий:
  - Теперь схема базы данных должна обновляться со всеми новыми таблицами, атрибутами и отношениями
  - Для каждого нового модуля должна быть установлена другая база данных.
  - Для двух новых модулей будет установлена новая база данных.
  - Новые модули должны соответствовать логической структуре текущей базы данных. **Правильный**
3. Одной из основных причин внедрения компаниями ERP является модульность. Могут ли компании закупать модули от разных производителей? Объясните свой **ответ**.

### **Внедрение и настройка ERP и других основных MIS. Процесс адаптации**

1. Затраты на внедрение ERP, связанные с внутренним персоналом, включают не только людей, которые составляют команды внедрения, но и время менеджеров, посвященных этому процессу.
  - Правда
  - Ложь
2. Внедрение ERP можно рассматривать как возможность *(выберите все подходящие варианты)*
  - Улучшите интеграцию данных внутри компании (и избегайте избыточности и несогласованности).
  - Принять новые технологические проекты, признанные лучшими в отрасли.
  - Позвольте сотрудникам решить, как выполнять свои задачи, чтобы каждый сотрудник мог настроить бизнес-процессы, в которых они участвуют.
3. После внедрения ERP функциональный менеджер жалуется, что новая система не

поддерживает способ выполнения определенных задач в рамках бизнес-процесса  
 Команда внедрения ERP обращается на горячую линию поддержки поставщика услуг ERP, и они подтверждают, что выполнение желаемых задач поддерживаться пакетом ERP (он предусмотрен в рамках его встроеной гибкости) . Какие существуют варианты решения проблемы?

### Информационные системы в менеджменте в процессе принятия решений

1. DSS - это:

- Система планирования принятия решений.
- Система поддержки диагностики.
- Система настройки данных.
- Система поддержки принятия решений.

2. Выделите признаки описывающие системы поддержки принятия решений:

(выберите все подходящие варианты)

- Данные из ERP отклоняются.
- Поддерживаются только неструктурные решения.
- Неструктурированные данные не сохраняются.
- Необходимо учитывать данные отдельных бизнес-операций

3. Могут ли у компании быть разные типы ИС, ориентированные на принятие решений, в зависимости от решений, которые они должны поддерживать, и данных, которые требуют эти решения.

Электронная таблица - это:

1. прикладная программа, предназначенная для обработки структурированных в виде таблицы данных;
2. прикладная программа для обработки кодовых таблиц;
3. устройство ПК, управляющее его ресурсами в процессе обработки данных в табличной форме;
4. системная программа, управляющая ресурсами ПК при обработке таблиц.

2. Выражение  $5(A_2+C_3):3(2B_2-3D_3)$  в электронной таблице имеет вид:

1.  $5(A_2+C_3)/3(2B_2-3D_3)$ ;
2.  $5*(A_2+C_3)/3*(2*B_2-3*D_3)$ ;
3.  $5*(A_2+C_3)/(3*(2*B_2-3*D_3))$ ;
4.  $5(A_2+C_3)/(3(2B_2-3D_3))$ .

3. Чему будет равно значение ячейки C1, если в нее ввести формулу =A1+B1:

	А	В	С
1	20	=A1/2	

1. 20;
2. 15;
3. 30;
4. 10?

4. Электронная таблица предназначена для:

1. обработки преимущественно числовых данных, структурированных с помощью таблиц;
2. упорядоченного хранения и обработки значительных массивов данных;
3. визуализации структурных связей между данными, представленными в таблицах;
4. редактирования графических представлений больших объемов информации.

5. При перемещении или копировании в электронной таблице абсолютные ссылки:

1. не изменяются;
2. преобразуются вне зависимости от нового положения формулы;
3. преобразуются в зависимости от нового положения формулы;
4. преобразуются в зависимости от длины формулы;

5. преобразуются в зависимости от правил указанных в формуле.
6. Адрес ячейки электронной таблицы – это
1. любая последовательность символов;
  2. номер байта оперативной памяти, отведенного под ячейку;
  3. имя, состоящее из имени столбца и номера строки;
  4. адрес байта оперативной памяти, отведенного под ячейку;
  5. адрес машинного слова оперативной памяти, отведенного под ячейку.
7. Чему будет равно значение ячейки A8, если в нее ввести формулу =СУММ(A1:A7)/2:
1. 280;

	A	B
1	10	
2	20	
3	30	
4	40	
5	50	
6	60	
7	70	
8	=СУММ(A1:A7)/2	
9		

2. 40;
3. 140;
4. 35?

## V. РЕСУРСЫ

### 5.1. Основная литература

1. Золотухина, Е.Б. Управление жизненным циклом информационных систем (продвинутый курс) [Электронный ресурс]: конспект лекций / Е.Б.Золотухина, С.А.Красникова, А.С.Вишня; ЭБС znanium. - М.:КУРС, НИЦ ИНФРА-М, 2017. - 119 с.: - Режим доступа: <http://znanium.com/bookread2.php?book=767219> . – Загл. с экрана.
2. Козлов, А.Ю. Статистический анализ данных в MS Excel [Электронный ресурс]: учебное пособие / А.Ю.Козлов, В.С.Мхитарян, В.Ф.Шишов; ЭБС znanium. - М.:НИЦ ИНФРА-М, 2016. - 320 с. - Режим доступа: <http://znanium.com/bookread2.php?book=558444> . – Загл. с экрана.
3. Форман, Дж. Много цифр. Анализ больших данных при помощи Excel [Электронный ресурс] / Джон Форман; пер. с англ.; БД AlpinaDigital. – М.: Альпина Паблшер, 2016. – 464 с. – ISBN 978-5-9614-5032-3. – Режим доступа: <http://lib.alpinadigital.ru/reader/book/7921>. – Загл. с экрана.

### 5.2. Дополнительная литература

1. Chaki, S. Enterprise Information Management in Practice: Managing Data and Leveraging Profits in Today's Complex Business Environment / Saumya Chaki; БД Books 24x7. – Apress, 2015. – 206 p. – ISBN 9781484212196. – Режим доступа: <https://www.books24x7.com/marc.asp?bookid=106738> . – Загл. с экрана.
2. Chessell, M., Smith, H.C. Patterns of Information Management / Mandy Chessell, Harald C. Smith; БД Books 24x7. – Kogan Page, 2013. – 733 p. – ISBN 9780133155501. – Режим доступа: <https://www.books24x7.com/marc.asp?bookid=57320> . – Загл. с экрана.
3. Doom, C. Introduction To Business Information Management / edited by Claude Doom; БД Books 24x7. – 2nd. ed. – ASP, 2013. – 191 p. – ISBN 9780133155501. – Режим доступа: <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=3115808>. – Загл. с экрана.
4. Kim, G. The Phoenix Project: A Novel About IT, DevOps, and Helping Your Business Win [Электронный ресурс] / Gene Kim, Kevin Behr, George Spafford; БД Books 24x7. – IT Revolution Press, 2018. – 432 с. – ISBN 978-1-942788-30-0. – Режим доступа: <https://www.books24x7.com/marc.asp?bookid=142715>. - Загл с экрана

5. Marketing Enterprise Information Management: When Information Becomes Inspiration / edited by P.Baan; БД Books 24x7. – Springer, 2013. – 234 p. – ISBN 9781461452355. – Режим доступа: <https://www.books24x7.com/marc.asp?bookid=76675>. – Загл. с экрана.
6. Гринберг, А.С. Информационный менеджмент [Электронный ресурс]: учебное пособие для вузов / А.С.Гринберг, В.Д.Маркова, С.А.Кузнецова; ЭБС Znanium. – М.: ЮНИТИ-ДАНА, 2015. – 415 с. – (Серия «Профессиональный учебник: Информатика»). – ISBN 5-238-00614-4. – Режим доступа: <http://znanium.com/bookread2.php?book=872151>. – Загл. с экрана.

### 5.3. Программное обеспечение

№ п/п	Наименование	Условия доступа
1	Windows Professional 8.1 Russian	<i>Из внутренней сети университета (договор)</i>
2	MS Office 2007 Prof +	<i>Из внутренней сети университета (договор)</i>

### 5.4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№ п/п	Наименование	Условия доступа/скачивания
1	Электронные образовательные ресурсы	<i>Договор на использование электронных баз данных/по подключению и обеспечению доступа к базам данных</i>

### 5.5. Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.