

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Кудрявцев Максим Геннадьевич  
Должность: Проректор по образовательной деятельности  
Дата подписания: 27.06.2022 в 10:58:56  
Уникальный программный ключ:  
790a1a8df2525774421adc1f96453f0e907bfb0

Министерство сельского хозяйства Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Российский государственный аграрный заочный университет»

Кафедра Экономики и финансов

Принято Ученым Советом  
ФГБОУ ВО РГАЗУ  
«26» января 2022 г. Протокол №9

«УТВЕРЖДЕНО»  
Проректор по образовательной  
деятельности М.А. Реньш  
«26» января 2022 г.



## Рабочая программа дисциплины

### Анализ данных

Направление подготовки 38.04.05 Бизнес-информатика

Направленность (профиль) программы Бизнес-аналитика

Квалификация магистр

Форма обучения очная

Балашиха 2022 г.

Рабочая программа разработана в соответствии с ФГОС ВО по направлению подготовки 38.04.05  
Бизнес-информатика

Рабочая программа дисциплины разработана профессором кафедры Экономики и финансов, д.э.н.,  
Аскеровым П. Ф.

Рецензент: профессор кафедры Управления, д.э.н. Васильева И. В.

# 1 Планируемые результаты обучения по дисциплине, соотнесенные с установленными в ОПОП ВО индикаторами достижения компетенций

## 1.1 Перечень компетенций, формируемых учебной дисциплиной

Универсальная компетенция	
ПК-2. Способен осуществлять сбор информации о состоянии аналитических работ в проекте, анализировать соответствия фактического состояния работ плановому состоянию.	<p>Знать (З): основы теории построения информационного общества; принципы построения современного информационного общества; сущность, основные функции, модели развития информационного общества; основные этапы и проблемы формирования информационного общества; проблемы становления информационного общества и деятельность государства по их решению; концептуальные основы формирования информационного общества в России.</p> <p>Уметь (У): применять новые информационно-коммуникационные технологии, которые являются базой становления информационного общества.</p> <p>Владеть (В): использованием вычислительной техники и базового программного и математического обеспечения в экономической (и других) предметных областях, создающих базисную структуру для дальнейшей профессиональной деятельности.</p>

## 2. Цели и задачи освоения учебной дисциплины, место дисциплины в структуре ОПОП ВО

Дисциплина «Анализ данных» относится к части, формируемой участниками образовательных отношений образовательной программы высшего образования 38.04.05 Бизнес-информатика, программа «Бизнес-аналитика».

**Цель:** формирование у базовых теоретических знаний в области теории вероятностей и математической статистики и способности к применению технологий обработки данных (в том числе big data) и машинного обучения к решению прикладных задач, связанных с оказанием финансовых услуг.

### Задачи:

- Овладение основными методами статистического анализа данных и математического моделирования применительно к социально-экономическим задачам.
- Выработка умения применять статистические методы для анализа информационных потоков с целью автоматизации задач принятия решений.

## 3. Объем учебной дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий, текущий и промежуточный контроль по дисциплине) и на самостоятельную работу обучающихся

### 3.1 Очная форма обучения

Вид учебной работы	4 семестр
--------------------	-----------

Общая трудоемкость дисциплины, зачетных единиц	3
<b>часов</b>	<b>108</b>
<b>Аудиторная (контактная) работа, часов</b>	<b>20,25</b>
в т.ч. занятия лекционного типа	10
занятия семинарского типа	10
промежуточная аттестация	0,25
<b>Самостоятельная работа обучающихся, часов</b>	<b>87,75</b>
Вид промежуточной аттестации	зачет

#### 4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

##### 4.1 Перечень разделов дисциплины с указанием трудоемкости аудиторной (контактной) и самостоятельной работы, видов контролей и перечня компетенций

Очная форма обучения

Наименование разделов и тем	Трудоемкость, часов			Наименование оценочного средства	Код компетенции
	всего	в том числе			
		аудиторной (контактной) работы	самостоятельной работы		
<b>Раздел 1. Анализ данных: методы, модели, инструменты</b>	50	10	40	Тест, реферат	ПК-2
1.1. Инфраструктура и управление данными	24	4	20		
1.2. Введение в теорию машинного обучения	26	6	20		
<b>Раздел 2. Введение в SAS: data step, основные процедуры</b>	59,75	20	39,75	Тест, реферат	ПК-2
2.1. Аналитические задачи в финансовом управлении	32,75	10	22,75		
2.2. Моделирование временных рядов. Стохастическое оптимизационное моделирование.	27	10	17		
<b>Итого за семестр</b>	104,75	30	105,75		
<b>Промежуточная аттестация</b>	4,25	0,25	4	Ответ на вопрос к зачету, итоговое тестирование	ПК-2
<b>ИТОГО по дисциплине</b>	108	20,25	87,75		

##### 4.2 Содержание дисциплины по разделам

###### Раздел 1. Анализ данных: методы, модели, инструменты

Цель: формирование теоретических знаний и практических навыков у обучающихся в области теории вероятности и характеристики распределения случайных величин, ограничения методов анализа данных, прикладных задач анализа данных и применении анализа данных в кредитном процессе банка: кредитная фабрика.

**Задачи:** изучить принципы и ключевые элементы стека Инфраструктура лаборатории анализа данных. Управление проектами в области анализа данных

**Перечень учебных элементов раздела:**

- 1.1. Инфраструктура и управление данными
- 1.2. Введение в теорию машинного обучения

**Раздел 2. Введение в SAS: data step, основные процедуры**

**Цель:** получить знания об основных понятиях, организации данных, организация системы, Data step и основных процедурах проведения.

**Задачи:** изучить прогнозные, оптимизационные и имитационные задачи в финансовом управлении и инструменты их решения

**Перечень учебных элементов раздела:**

- 2.1. Аналитические задачи в финансовом управлении
- 2.2. Моделирование временных рядов. Стохастическое оптимизационное моделирование.

**5. Оценочные материалы по дисциплине**

Оценочные материалы по дисциплине представлены в виде фонда оценочных средств.

**6. Материально-техническое и учебно-методическое обеспечение дисциплины**

**6.1 Перечень учебно-методического обеспечения по дисциплине**

№ п/п	Автор, название, место издания, издательство, год издания, количество страниц, режим доступа
1	Методические указания по изучению дисциплины

**6.2 Перечень учебных изданий, необходимых для освоения дисциплины \***

Электронные учебные издания в электронно-библиотечных системах (ЭБС)\*\*:

№ п/п	Автор, название, место издания, год издания, количество страниц	Ссылка на учебное издание в ЭБС
<b>Основная:</b>		
1	Ковалева М. А., Волошин С. Б. Анализ данных. Учебное пособие – М.: Мир науки, 2019 – Сетевое издание. Режим доступа:	<a href="https://izd-mn.com/PDF/32MNNPU19.pdf">https://izd-mn.com/PDF/32MNNPU19.pdf</a>
2	Белов В.С. Информационно-аналитические системы. Основы проектирования и применения. Учебное пособие. - М: МЭСИ, 2004.	<a href="http://shpora1.do.am/_ld/2/255.pdf">http://shpora1.do.am/_ld/2/255.pdf</a>
<b>Дополнительная</b>		
1	Большакова Е.И., Груздева Н.В. Основы программирования на языке Лисп: Учебное пособие. - М.: МАКС Пресс, 2010	<a href="http://www.recyclebin.ru/BMK/LISP/lisp.html">http://www.recyclebin.ru/BMK/LISP/lisp.html</a>

2	Ласковец С.В. Методология научного творчества [Электронный ресурс]: Учебное пособие. - Москва: Евразийский открытый институт, 2010. - 32 с]	<a href="http://www.biblioclub.ru/90384_Metodologiya_nauchnogo_tvorchestva_Uchebnoe_posobie.html">http://www.biblioclub.ru/90384_Metodologiya_nauchnogo_tvorchestva_Uchebnoe_posobie.html</a>
---	---	---

### 6.3 Перечень электронных образовательных ресурсов \*

№ п/п	Электронный образовательный ресурс	Доступ в ЭОР (сеть Интернет, локальная сеть, авторизованный/свободный доступ)
1	Образовательная платформа Coursera. [Электронный ресурс]. - Режим доступа: - Загл. с экрана	<a href="https://www.coursera.org/">https://www.coursera.org/</a>
2	MachineLearning.ru	<a href="http://machinelearning.ru">http://machinelearning.ru</a>
3	ProQuest [Электронный ресурс]. – URL:	<a href="http://search.proquest.com/index">http://search.proquest.com/index</a>
4	Научная электронная библиотека «КиберЛенинка» [Электронный ресурс].	<a href="http://cyberleninka.ru//">http://cyberleninka.ru//</a>

### 6.4 Современные профессиональные базы данных, информационные справочные системы и лицензионное программное обеспечение

#### Современные профессиональные базы данных

<https://rosstat.gov.ru/> - Федеральная служба государственной статистики.

<https://cyberleninka.ru/> - научная электронная библиотека открытого доступа (Open Access).

<http://link.springer.com/> - полнотекстовая коллекция (база данных) электронных книг издательства Springer Nature.

<http://fcior.edu.ru/> - Федеральный центр информационно-образовательных ресурсов.

<https://agris.fao.org/agris-search/index.do> - Международная информационная система по сельскохозяйственным наукам и технологиям.

<http://window.edu.ru/> - Информационная система «Единое окно доступа к образовательным ресурсам»

#### Информационные справочные системы

1. Информационно-справочная система «Гарант». – URL: <https://www.garant.ru/>

2. Информационно-справочная система «Консультант Плюс». – URL: <http://www.consultant.ru/>

3. Федеральный образовательный портал "Информационно-коммуникационные технологии в образовании". – URL: <http://www.ict.edu.ru>

#### Лицензионное программное обеспечение

Microsoft Office (Access, Excel, PowerPoint, Word и т. д),

OpenOffice, Люникс (бесплатное программное обеспечение широкого класса),

система дистанционного обучения Moodle ([www.edu.rgazu.ru](http://www.edu.rgazu.ru)),

Вебинар (Adobe Connect v.8, Zomm, Google Meet, Skype, Мираполис), программное обеспечение электронного ресурса сайта, включая ЭБС AgriLib и видеоканал РГАЗУ (<http://www.youtube.com/rgazu>),

антивирусное программное обеспечение Dr. WEB Desktop Security Suite.

### 6.5 Перечень учебных аудиторий, оборудования и технических средств обучения\*\*

Предназначение помещения (аудитории)	Наименование корпуса, № помещения (аудитории)	Перечень оборудования (в т.ч. виртуальные аналоги) и технических средств обучения*

Для занятий лекционного типа	Учебно-административный корпус. Каб. 135. Учебная аудитория для проведения лекционных занятий (поточная).	Специализированная мебель, набор демонстрационного оборудования и учебно-наглядных пособий, проектор EPSON EB-1880, экран настенный моторизованный SimSCREEN
Для занятий семинарского типа, групповых консультаций, промежуточной аттестации	Учебно-административный корпус. Каб. 142	Специализированная мебель, набор демонстрационного оборудования, персональные компьютеры 14 шт. на базе ASUSP5KPL-CM/2048 RAM/DDR2/Intel Core 2Duo E7500, 2,9 MHz/AtiRadeon HD 4350 512 Mb/HDD 250/Win7-32/MSOffice 2010/Acer V203H
Для самостоятельной работы	Учебно-административный корпус. Помещение для самостоятельной работы. Читальный зал	Персональные компьютеры 11 шт. Выход в интернет, доступ в электронную информационно-образовательную среду университета
	Учебно-лабораторный корпус. Помещение для самостоятельной работы. Каб. 320.	Специализированная мебель, набор демонстрационного оборудования, персональные компьютеры 11 шт. Выход в интернет, доступ в электронную информационно-образовательную среду университета
	Учебно-административный корпус. Каб. 105. Учебная аудитория для учебных занятий обучающихся из числа инвалидов и лиц с ОВЗ	Специализированная мебель, набор демонстрационного оборудования. Автоматизированное рабочее место для инвалидов-колясочников с коррекционной техникой и индукционной системой ЭлСис 290; Автоматизированное рабочее место для слабовидящих и незрячих пользователей со стационарным видеоувеличителем ЭлСис 29 ON; Автоматизированное рабочее место для слабовидящих и незрячих пользователей с портативным видеоувеличителем ЭлСис 207 CF; Автоматизированное рабочее место для слабовидящих и незрячих пользователей с читающей машиной ЭлСис 207 CN; Аппаратный комплекс с функцией видеоувеличения и чтения для слабовидящих и незрячих пользователей ЭлСис 207 OS.

**Министерство сельского хозяйства Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Российский государственный аграрный заочный университет»**

**Фонд оценочных средств для проведения текущего контроля и промежуточной  
аттестации обучающихся по дисциплине  
Анализ данных**

Направление подготовки 38.04.05 Бизнес-информатика

Направленность (профиль) программы Бизнес-аналитика

Квалификация магистр

Форма обучения очная

Балашиха 2022 г.



### 1. Описание показателей и критериев оценивания планируемых результатов обучения по учебной дисциплине

Компетенций	Индикатор сформированности компетенций	Уровень освоения*	Планируемые результаты обучения	Наименование оценочного средства
<p>ПК-2. Способен осуществлять сбор информации о состоянии аналитических работ в проекте, анализировать соответствия фактического состояния плановому состоянию.</p>	<p>Знать (З): основы теории построения информационного общества; принципы построения современного информационного общества; сущность, основные функции, модели развития информационного общества; основные этапы и проблемы формирования информационного общества; проблемы становления информационного общества и деятельность государства по их решению; концептуальные основы формирования информационного общества в России.</p>	<p><b>Пороговый (удовлетворительно)</b></p>	<p><b>знать:</b> основы теории построения информационного общества; принципы построения современного информационного общества; сущность, основные функции, модели развития информационного общества; основные этапы и проблемы формирования информационного общества; проблемы становления информационного общества и деятельность государства по их решению; концептуальные основы формирования информационного общества в России.</p> <p><b>уметь:</b> применять новые информационно-коммуникационные технологии, которые являются базой становления информационного общества.</p> <p><b>владеть:</b> использованием вычислительной техники и базового программного и математического обеспечения в экономической (и других) предметных областях, создающих базисную структуру для дальнейшей профессиональной деятельности</p>	<p>Тест, реферат, ответ на вопрос к зачету, выполнение итогового тестирования</p>
	<p>Уметь (У): применять новые информационно-коммуникационные технологии, которые являются базой становления информационного общества.</p> <p>Владеть (В): использованием вычислительной техники и базового программного и математического обеспечения в</p>		<p><b>Продвинутый (хорошо)</b></p>	<p><b>Знает твердо:</b> основы теории построения информационного общества; принципы построения современного информационного общества; сущность, основные функции, модели развития информационного общества; основные этапы и проблемы формирования информационного общества; проблемы становления информационного общества и деятельность государства по их решению; концептуальные основы формирования информационного общества в России.</p> <p><b>Умеет уверенно:</b> применять новые информационно-коммуникационные технологии, которые являются базой</p>

	экономической (и других) предметных областях, создающих базисную структуру для дальнейшей профессиональной деятельности.		становления информационного общества. <b>Владет уверенно:</b> использованием вычислительной техники и базового программного и математического обеспечения в экономической (и других) предметных областях, создающих базисную структуру для дальнейшей профессиональной деятельности.	
		<b>Высокий (отлично)</b>	<b>Имеет сформировавшееся систематические знания:</b> об основах теории построения информационного общества; принципы построения современного информационного общества; сущность, основные функции, модели развития информационного общества; основные этапы и проблемы формирования информационного общества; проблемы становления информационного общества и деятельность государства по их решению; концептуальные основы формирования информационного общества в России. <b>Имеет сформировавшееся систематическое умение:</b> применять новые информационно-коммуникационные технологии, которые являются базой становления информационного общества. <b>Показал сформировавшееся систематическое владение:</b> использованием вычислительной техники и базового программного и математического обеспечения в экономической (и других) предметных областях, создающих базисную структуру для дальнейшей профессиональной деятельности.	Тест, реферат, ответ на вопрос к зачету, выполнение итогового тестирования

## 2. Описание шкал оценивания

### 2.1 Шкала оценивания на этапе текущего контроля

Форма текущего контроля	Отсутствие усвоения (ниже порогового)*	Пороговый (удовлетворительно)	Продвинутый (хорошо)	Высокий (отлично)
Тест	Менее 51%	51-79%	80-90%	91% и более
Реферат	не выполнена или все	Цель и задачи реферата	Цель и задачи выполнения	Цель написания реферата

	задания решены неправильно	достигнуты частично. Актуальность темы реферата определена неубедительно. В реферате выявлены значительные отклонения от требований методических указаний.	реферата достигнуты. Актуальность темы реферата подтверждена. Реферат выполнен с незначительными отклонениями от требований методических указаний.	достигнута, задачи решены. Актуальность темы исследования корректно и полно обоснована. Реферат выполнен согласно требованиям.
--	-------------------------------	---	---	--

### 2.2 Шкала оценивания на этапе промежуточной аттестации (экзамен)

Форма промежуточной аттестации	Отсутствие усвоения (ниже порогового)	Пороговый (удовлетворительно)	Продвинутый (хорошо)	Высокий (отлично)
Ответ на вопрос к зачету	не выполнена или все задания решены неправильно	Цель и задачи вопроса достигнуты частично.	Цель и задачи выполнения вопроса достигнуты.	Цель написания ответа на вопрос достигнута, задачи решены.
Выполнение итогового тестирования	Менее 51%	51-79%	80-90%	91% и более

- 3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ  
КОМПЛЕКТ ТЕСТОВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ**

**Раздел 1. Анализ данных: методы, модели, инструменты.**

**1. Суммарная оперативная память IBM Watson составляет порядка:**

- А) 100 гигабайт
- Б) 5000 терабайт
- В) 10 зетабайт
- Г) 15 терабайт

**2. Кто ввел термин Большие данные?**

- А) Клиффорд Линч
- Б) Алан Тьюринг
- В) Бьерн Страуструп
- Г) Дональд Кнут

**3. Какие данные занимают больше мировой памяти относительно остальных?**

- А) Structured Data
- Б) Unstructured Data
- В) Semi-Structured Data
- Г) Quasi-Structured Data

**4. BigData – это ...**

- А) Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации, или обработки.
- Б) Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.
- В) Колоссальный объем данных, собранных человечеством.
- Г) Класс в Java, предназначенный для хранения данных от 100 Гб

**5. Какая компания создала технологию MapReduce?**

- А) Google
- Б) Yahoo
- В) EMC
- Г) Oracle

**6. Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:**

- А) Структурированными
- Б) Полуструктурированными
- В) Квазиструктурированными
- Г) Неструктурированными

**7. Что означает термин «Big Data» в информационных технологиях?**

- А) Комплексный набор методов для создания файлов большого объема
- Б) Комплексный набор методов обработки структурированных и

неструктурированных данных колоссальных объемов.

В) Файлы с большим количеством данных.

Г) Представление времени, дня, месяца и года в качестве значения количества миллисекунд, прошедших с начала нашей эры.

**8. Данные имеющие определенный тип, формат и структуру (например, транзакционные данные) являются:**

А) Структурированными

Б) Полуструктурированными

В) Квазиструктурированными

Г) Неструктурированными

**9. Чему примерно равен объем всей существующей на земле информации (в байтах)?**

А)  $10^{11}$

Б)  $10^{21}$

В)  $10^{1010101}$

Г)  $10^{171}$

**10. В каком году впервые был введен термин Большие данные?**

А) 2002

Б) 2004

В) 2006

Г) 2008

## **Раздел 2. Введение в SAS: data step, основные процедуры.**

**1. Что является средством анализа в BI?**

А) Карты показателей;

Б) Совместная работа и управление рабочими процессами;

В) Информационные панели;

Г) BI инфраструктура.

**2. Основное умение исследователя данных?**

А) Умение находить наиболее важные элементы в хранимой информации

Б) Уметь прогнозировать исход работы системы

В) Находить скрытые логические связи в системе собранной информации

Г) Отличать неструктурированные данные от структурированных

**3. Какой язык программирования из перечисленных является наиболее важным для аналитика?**

А) C++

Б) PHP

В) F#

Г) R

**4. Что означает термин «Business Intelligence» в информационных технологиях?**

А) Комплексный набор методов для создания бизнес планов.

Б) Методы и инструменты для перевода необработанной информации в осмысленную, удобную для восприятия форму.

В) Файлы, содержащие информацию о бизнес плане.

Г) Технологии, направленные на развитие бизнеса.

**5. Языком, на котором был разработан RabbitMQ, является:**

- A) Java
- Б) Python
- В) C++
- Г) Erlang

**6. Что является главным результатом процесса Business Intelligence?**

- A) Возможность принятия решений для бизнеса
- Б) Результаты интеллектуального анализа данных
- В) Возможность использования искусственного интеллекта
- Г) Получение структуризации данных после выполнения всех шагов процесса

**7. Что из перечисленного не является средством анализа?**

- A) Продвинутая визуализация
- Б) Reporting
- В) Predictive Modelling
- Г) Data Mining

**8. Что относится к средствам предоставления информации в «Business Intelligence»?**

- A) Генератор нерегламентированных запросов
- Б) Совместная работа и управление рабочими процессами
- В) Предиктивное моделирование и Data Mining
- Г) Карты показателей

**9. Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является:**

- A) OLAP
- Б) Data Mining
- В) Predictive Modelling
- Г) Data Science

**10. Что не является целью процесса Business Intelligence?**

- A) Интерпретация большого количества данных;
- Б) Моделирование исходов различных вариантов действий;
- В) Модификация существующего программного обеспечения;
- Г) Отслеживание результатов решений.

## **ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ**

### **ПРИМЕРНЫЕ ТЕМЫ РЕФЕРАТОВ**

#### **Раздел 1. Анализ данных: методы, модели, инструменты.**

1. Современные тенденции развития Big Data в мировой компьютерной индустрии.
2. Современные тенденции развития Big Data в Российской Федерации.
3. Предпосылки развития Big Data.
4. Современное состояние операционной системы Hadoop.
5. Отрасли народного хозяйства, в которых может быть задействована операционная система Hadoop.

#### **Раздел 2. Введение в SAS: data step, основные процедуры**

1. Перечень народно-хозяйственных задач, которые целесообразно решать с помощью технологии Map Reduce.
2. Применение технологий NoSQL в современных бизнес проектах.
3. Перспективы применения среды R для анализа статистических данных, необходимых для принятия решений в управлении.
4. Отрасли народного хозяйства, в которых могут быть задействованы системы класса Data Mining.
5. Перспективы применения кластерного анализа и его результатов для управления экономикой.

## КОМПЛЕКТ ПРИМЕРНЫХ ВОПРОСОВ (зачет) ПО ДИСЦИПЛИНЕ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

### Примерный перечень вопросов к зачету

1. Термин Big Data, цель и главные характеристики больших данных.
2. Принцип трех "V".
3. Структурированные, неструктурированные и квази-структурированные данные.
4. Business Intelligence.
5. Цели и задачи Data Science.
6. Что такое жизненный цикл аналитики данных.
7. Технология Apache Hadoop.
8. Технология MapReduce, достоинства и недостатки.
9. Файловая система HDFS.
10. Хранилища NoSQL главные отличия от обычных БД.
11. Теорема CAP и ее следствия.
12. Amazon S3.
13. Язык R.
14. Основные задачи, требования и типы визуализации.
15. Data Mining и основные типы.
16. Классификация и кластеризация
17. Современные инструменты анализа данных, их типы и задачи.
18. Основные метрики оценки взаимосвязанности между случайными величинами.
19. Ключевые элементы стека Инфраструктура лаборатории анализа данных.
20. Hadoop: верхне-уровневая архитектура, ключевые принципы работы
21. Архитектура кредитной фабрики.
22. Понятия DevOps, MLOps.
23. Основные принципы организации процесса разработки и внедрения моделей.
24. Основные направления применения моделей, применяемые в рамках кредитной фабрики.
25. Подходы к оценке качества моделей, понятия переобученной и недоучившейся моделей.
26. Подходы к нормализации, трансформации, очистке и обогащению данных.
27. Основные подходы feature engineering
28. SAS data step, основные процедуры SAS language.
29. Методы прогнозирования временных рядов.
30. Применение методов оптимизации в рамках задач финансового управления.

## КОМПЛЕКТ ПРИМЕРНЫХ ТЕСТОВЫХ ЗАДАНИЙ (зачета) ПО ДИСЦИПЛИНЕ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

### 1. Что из этого не является реализацией Hadoop?

- А) Google MapReduce
- Б) Phoenix
- В) GreenMint
- Г) Qizmt

### 2. Какие из перечисленных пунктов являются достоинствами MapReduce?

- А) Оптимальная производительность
- Б) Эффективное применение в маленьких кластерах с небольшим объемом данных
- В) Масштабируемость
- Г) Отказоустойчивость

### 3. Что такое Oozie?

- А) Распределенный координационный сервис
- Б) Нереляционная распределенная база данных
- В) Язык управления потоком данных и исполнительная среда для анализа больших объемов данных
- Г) Сервис для записи и планировки заданий Hadoop

**4. Сколько уровней имеет лямбда-архитектура?**

- А) 2
- Б) 3
- В) 4
- Г) 5

**5. Какие компоненты являются частями MapReduce?**

- А) Task Tracker
- Б) Name Node и Data Node
- В) Job Tracker и Task Tracker
- Г) Job Tracker, Task Tracker, Name Node и Data Node

**6. Что такое Spark?**

- А) Инструмент для кластерных вычислений
- Б) Графический движок
- В) Библиотека для работы с графами
- Г) Технология распределенных вычислений

**7. Дайте определение Map Reduce...**

- А) Модель распределенных вычислений, предназначенная для параллельных вычислений над очень большими (до нескольких петабайт) объемами данных
- Б) Набор компонентов и интерфейсов для распределенных файловых систем и общего ввода-вывода
- В) Распределенная файловая система, работающая на больших кластерах типовых машин
- Г) Распределенный сервис для коллекционирования, сбора, и перемещения больших массивов данных

**8. Что из этого является недостатком MapReduce?**

- А) Фиксированный алгоритм обработки данных
- Б) Масштабируемость
- В) Отказоустойчивость
- Г) Возможность автоматического распараллеливания

**9. Какое API было добавлено в Hadoop v2.0?**

- А) YAWN
- Б) YARN
- В) SARN
- Г) DARN

**10. Какая цель у NameNode в HDFS?**

- А) Хранить индекс того, какая часть данных находится в каком узле
- Б) Хранить имя файла, хранящегося в конкретном узле
- В) Хранить индекс узла, в котором хранится имя файла
- Г) Хранить имена узлов

**11. Вертикальное масштабирование...**

- А) Требует изменений в прикладных программах, работающих на таких системах
- Б) Не требует никаких изменений в прикладных программах, работающих на таких системах
- В) Уменьшает производительность каждого компонента БД
- Г) Увеличивает скорость загрузки данных

**12. Для достижения какого свойства в БД типа NoSQL нет JOIN операций?**

- А) Intercepting
- Б) Concurrency



- В) Consistency
- Г) Capacity

**13. Что, согласно теореме CAP (теореме Брюера), возможно обеспечить в любой реализации распределённых вычислений?**

- А) Только согласованность данных
- Б) Только доступность данных
- В) Согласованность данных, доступность данных, устойчивость к разделению
- Г) Не более двух свойств из трёх вышеприведённых

**14. Выберите верное определение понятия AP-система:**

- А) Система, во всех узлах которой данные согласованы и обеспечена доступность, жертвует устойчивостью к распаду на секции
- Б) Распределённая система, в каждый момент обеспечивающая целостный результат и способная функционировать в условиях распада
- В) Распределённая система, отказывающаяся от целостности результата
- Г) Система, автоматически изменяющая данные алгоритма своего с целью сохранения оптимального состояния

**15. Что означает термин NoSQL?**

- А) Не SQL
- Б) Не только SQL
- В) Без SQL
- Г) SQL – плохо

**16. Разбиение системы на более мелкие структурные компоненты и разнесение их по отдельным физическим машинам (или их группам), и (или) увеличение количества серверов, параллельно выполняющих одну и ту же функцию, это:**

- А) Горизонтальное масштабирование
- Б) Вертикальное масштабирование
- В) Master- slave репликация
- Г) Peer-to-peer репликация

**17. Что из перечисленного относится к графо-ориентированным хранилищам (Graph Store)?**

- А) Neo4j
- Б) BaseX
- В) Elasticsearch
- Г) Ничего

**18. Что поддерживает NoSQL?**

- А) Операцию Insert
- Б) Полностью стандарт SQL
- В) Операцию Join
- Г) Операцию Group by

**19. Какие три свойства фигурируют в определении теоремы CAP?**

- А) Согласованность данных
- Б) Сложность
- В) Доступность
- Г) Устойчивость к разделению

**20. Выделение таблицы или группы таблиц на отдельный сервер это...**

- А) Горизонтальное масштабирование
- Б) Вертикальное масштабирование
- В) Горизонтальный шардинг
- Г) Вертикальный шардинг