

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ  
ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
СТАВРОПОЛЬСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ  
УНИВЕРСИТЕТ**



Учетно-финансовый факультет  
Кафедра «Экономическая безопасность, статистика и эконометрика»

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ  
САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПО  
ДИСЦИПЛИНЕ  
«ТЕХНОЛОГИИ БОЛЬШИХ ДАННЫХ»  
по направлению 38.03.01 Экономика**

(код и наименование направления подготовки)

---

**Учетно-аналитическое обеспечение цифровой экономики**

---

## Комплект вопросов для собеседования

### **Тема 1. Предпосылки формирования тренда больших данных**

1. Основные вызовы больших данных (4V)
2. Определение термина «большие данные»
3. Базовое представление о Map Reduce и
4. Hadoop
5. Представление о работе аналитика

### **Тема 2. Инструменты для обработки больших данных**

1. Знакомство с языками и прикладными пакетами для обработки больших данных.
2. Рассмотрение общей концепции и синтаксиса языка R (примеры).

### **Тема 3. Аналитика больших данных**

1. Процесс аналитики
2. Стандарты жизненного цикла Big Data: CRISP-DM
3. Принципы и инструменты аналитики
4. Задачи и компетенции аналитиков Big Data
5. Big Data как рынок
6. Стек технологий
7. По поддержки принятия решений
8. Игроки на рынке BD
9. Крупнейшие проекты BD в России

### **Тема 4. Когнитивный анализ данных**

1. Введение в Data Mining – понятие, структура, составляющие и сопутствующие науки.
2. Задачи Data Mining и способы их решения. Классификация методов DM.
3. Области применения DM.
4. Классы систем DM.
5. Процесс накопления и анализа данных: Азбука когнитивного анализа.

### **Тема 5. Аналитика больших данных. Математическая статистика**

1. Шкалы измерений.
2. Генеральная совокупность и выборка.
3. Нормальное распределение. Уровень статистической достоверности.
4. Свойства описательных статистик (Дескриптивный анализ)
5. Визуальное представление данных
6. Меры изменчивости

### **Тема 6. Методы DATA MINING**

1. Данные & знания
2. Типовые задачи Data Mining
3. Обучаемые и необучаемые задачи
4. Жизненный цикл проекта DM
5. Математический аппарат DM
6. Стандарты DM

### **Тема 7. Прикладные инструменты анализа данных. Корреляция**

1. Готовые комплексные решения: Weka, RapidMiner, Knime, Orange IBM SPSS Modeler ( в прошлом Clementine )
2. Инструменты визуализации: Tableau,
3. Фреймворки на JS, D3
4. Корреляция. Понятие корреляции
5. Значимость коэффициента корреляции
6. Виды связи между переменными

## **Фонд кейс-задач для интерактивных практических занятий**

### **Кейс-задача для практического занятия №6**

#### **Тема: Методология оценки национальной экономической безопасности**

Этапы работы с кейсом.

#### 1. Этап введения в изучаемую проблему

По исходным данным о темпах экономического роста проведите статическую оценку индикатора национальной экономической безопасности. Для этого рассчитайте индивидуальные и средние относительные показатели сравнения (ОПС). Проведите сравнительную оценку уровней ОПС за 2017 год со средними ОПС за 2000-2017 гг. Сделайте выводы.

Год	Тем экономического роста, раз
2000	1,125
2001	1,075
2002	1,064
2003	1,095
2004	1,101
2005	1,098
2006	1,115
2007	1,114
2008	1,073
2009	0,929
2010	1,058
2011	1,072
2012	1,056
2013	1,034
2014	1,025
2015	0,986
2016	1,011
2017	1,034

#### 2. Анализ ситуации.

Подгруппы обучающихся (по 3-5 человек) проводят анализ представленной информации, совершают предварительные расчеты показателей на основании наиболее подходящих, по их мнению, методик оценки и качественной интерпретации показателей.

#### 3. Этап презентации.

Каждая подгруппа представляет альтернативные результаты исследования к обсуждению.

#### 4. Этап общей дискуссии.

Представленные альтернативные варианты решения обсуждаются в ходе общей дискуссии. Оцениваются преимущества и недостатки.

#### 5. Этап подведения итогов.

На основании приведенных решений делается качественная характеристика полученных результатов.

### Кейс-задача для практического занятия №12

#### Тема: Продовольственная безопасность

Этапы работы с кейсом.

#### 1. Этап введения в изучаемую проблему

В таблице 5.1 представлена динамика ресурсов и использования молока и молокопродуктов в 1995-2017 гг.

Таблица 5.1. – Ресурсы и использования молока и молокопродуктов, млн т

Год	Ресурсы				Использование				Запасы (КГ)
	Запасы (НГ)	Производство	Импорт	Итого	Производственное потребление	Потери	Экспорт	Личное потребление	
1995	1,8	39,2	6,3	47,3	7,0	0,10	0,4	37,4	2,4
1996	2,4	35,8	4,5	42,8	6,7	0,05	0,49	34,2	1,3
1997	1,3	34,1	6,3	41,7	6,4	0,04	0,2	33,7	1,4
1998	1,4	33,3	4,9	39,6	5,8	0,04	0,3	32,3	1,2
1999	1,2	32,3	4,7	38,2	5,3	0,03	0,2	31,3	1,3
2000	1,3	32,3	4,7	38,3	5,2	0,03	0,5	31,3	1,2
2001	1,2	32,9	4,9	39	5,2	0,03	0,6	31,8	1,4
2002	1,4	33,5	4,9	39,8	5,2	0,03	0,4	32,8	1,3
2003	1,4	33,3	5,6	40,3	5	0,02	0,5	33,1	1,7
2004	1,7	31,8	6,3	39,8	4,5	0,02	0,5	33,1	1,7
2005	1,7	30,8	7,1	39,6	4,1	0,02	0,5	33,2	1,8
2006	1,8	31,1	7,3	40,2	4,1	0,02	0,5	33,7	1,9
2007	1,9	32	7,1	41	4,2	0,02	0,6	34,3	1,9
2008	1,9	32,4	7,3	41,6	4,3	0,02	0,6	34,6	2,08
2009	2,1	32,6	7,0	41,7	4,4	0,02	0,5	34,9	1,88
2010	1,9	31,8	8,2	41,9	4,3	0,03	0,5	35,2	1,87
2011	1,9	31,7	7,9	41,5	3,7	0,03	0,6	35,2	1,97
2012	2,0	31,8	8,5	42,3	3,9	0,03	0,7	35,7	1,97
2013	2,0	30,7	9,4	42,1	3,8	0,02	0,6	35,8	1,88
2014	1,9	30,8	9,2	41,9	3,5	0,03	0,6	35,7	2,07
2015	2,1	30,8	7,9	40,8	3,3	0,03	0,6	34,9	2
2016	1,9	30,8	7,5	40,2	3,2	0,03	0,6	34,7	1,7
2017	1,7	31,1	6,6	39,4	2,9	0,03	0,6	34,2	1,7

Оцените цепные темпы роста (индексы динамики) каждого показателя.

Рассчитайте абсолютные показатели чистого экспорта.

Рассчитайте индексы независимости по молоку и молокопродуктам.

Рассчитайте рацион потребления молока и молокопродуктов.

Сделайте выводы.

#### 2. Анализ ситуации.

Подгруппы обучающихся (по 3-5 человек) проводят анализ представленной информации, совершают предварительные расчеты показателей на основании наиболее подходящих, по их мнению, методик оценки и качественной интерпретации показателей.

#### 3. Этап презентации.

Каждая подгруппа представляет альтернативные результаты исследования к обсуждению.

#### 4. Этап общей дискуссии.

Представленные альтернативные варианты решения обсуждаются в ходе общей дискуссии. Оцениваются преимущества и недостатки.

#### 5. Этап подведения итогов.

На основании приведенных решений делается качественная характеристика полученных результатов.

### **Вопросы к экзамену:**

1. Понятие Большие данные. Роль цифровой информации в 21 веке. Проблемы анализа и обработки большого объема данных.
2. Базовые принципы обработки больших данных.
3. Определение модели. Свойства модели.
4. Аналитический подход к моделированию.
5. Информационный подход к моделированию.
6. Лица, участвующие в информационном моделировании. Общая схема анализа.
7. Определение тиражирования знаний. Процесс построения модели.
8. Технологии обработки больших данных: NoSQL, MapReduce, Hadoop, R.
9. Методика извлечения знаний Knowledge Discovery in Databases (KDD). Этапы KDD.
10. Data Mining. Постановка основных задач.
11. Машинное обучение. Бизнес-решения с помощью алгоритмов Data Mining.
12. Классификация ПО в области Data Mining и KDD. Типовая схема системы на базе аналитической платформы.
13. Понятие ассоциативного правила и транзакции. Определение поддержки и достоверности. Определение значимости и полезности ассоциативных правил, показатели их характеризующие.
14. Формальная постановка задачи кластеризации. Цели кластеризации.
15. Основные шаги алгоритма k-means. Условие остановки алгоритма k-means. Преимущества и недостатки алгоритма k-means.
16. Кластеризация с помощью самоорганизующейся карты Кохонена
17. Этапы проведения классификации. Обзор методов классификации и регрессии.
18. Задачи линейной и логистической регрессии.
19. Определение дерева решений. Структура дерева решений. Выбор атрибута разбиения в узле.
20. Алгоритм ID3.  
Алгоритм C4.5.