

**Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА  
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ  
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

---

**Факультет Высшая школа финансов и менеджмента**

Кафедра финансового менеджмента, управленческого учета и международных  
стандартов финансовой деятельности

УТВЕРЖДЕНА

решением кафедры

Протокол №6 от «07» сентября 2017 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Б1.В.ДВ.03.01 «Информационные технологии в прогнозно-аналитической дея-  
тельности»**

направление подготовки  
38.03.02 Менеджмент

направленность (профиль)  
Финансовый менеджмент

квалификация: бакалавр  
формы обучения: очно-заочная

Год набора – 2018

Москва, 2017 г.

**Автор(ы)–составитель(и):**

преподаватель кафедры финансового менеджмента, управленческого учета и международных стандартов финансовой деятельности \_\_\_\_\_ Е.А. Баркова

Заведующий кафедрой финансового менеджмента, управленческого учета и международных стандартов финансовой деятельности

\_\_\_\_\_ д.э.н., профессор Е.Н. Лобанова

## СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы .....	4
2. Объем и место дисциплины в структуре ОП ВО .....	5
3. Содержание и структура дисциплины .....	5
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине .....	7
4.1. Формы и методы текущего контроля успеваемости и промежуточной аттестации .....	7
4.2. Материалы текущего контроля успеваемости .....	8
4.3. Оценочные средства для промежуточной аттестации .....	20
5. Методические указания для обучающихся по освоению дисциплины .....	28
6. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине .....	29
6.1 Основная литература .....	29
6.2 Дополнительная литература .....	29
6.3 Интернет-ресурсы .....	29
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы .....	30

# **1. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

1.1. Дисциплина Б1.В.ДВ.03.01 «Информационные технологии в прогнозно-аналитической деятельности» обеспечивает овладение следующими компетенциями с учетом этапа :

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ОПК-7	Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учётом основных требований информационной безопасности	ОПК-7.3	Способность решить стандартные задачи профессиональной деятельности на основе информационной культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности

1.2. В результате освоения дисциплины Б1.В.ДВ.03.01 «Информационные технологии в прогнозно-аналитической деятельности» у студентов должны быть:

ОТФ/ТФ (при наличии профстандарта)/ профессиональные действия	Код этапа освоения компетенции	Результаты обучения
Тактическое управление процессами планирования и организации производства	ОПК-7.3	Способен оценить необходимость и возможность применения интеллектуального анализа данных при решении конкретных бизнес-задач; Способен осуществлять выбор методов анализа, подготовку исходных данных, визуализацию и интерпретацию полученных результатов

## 2. ОБЪЕМ И МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина (модуль) «Информационные технологии в прогнозно-аналитической деятельности» относится к вариативной части Блока 1, является дисциплиной по выбору и изучается в 6 семестре. Код дисциплины Б1.В.ДВ.03.01

Освоение дисциплины опирается на минимально необходимый объем теоретических знаний в области математики, статистики и эконометрики.

Дисциплина «Информационные технологии в прогнозно-аналитической деятельности» реализуется после изучения следующих дисциплин: «Математика», «Статистика».

Общая трудоемкость дисциплины составляет 72 (2 ЗЕ).

Из них:

32 часа — контактная работа с преподавателем (компьютерный практикум);

40 часов — самостоятельная работа обучающихся.

Форма промежуточной аттестации — зачет.

## 3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Таблица 1

№ п/п	Наименование тем	Объем дисциплины (модуля), час.						Форма текущего контроля успеваемост и, промежуточ ной аттестации*
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Очная форма обучения								
Тема 1.	Аналитическая отчетность.	8			4		4	Т, КР
Тема 2.	Классификация и регрессия. Статистические методы.	12			6		6	Т
Тема 3.	Классификация и регрессия. Машинное обучение.	12			6		6	Т

№ п/п	Наименование тем	Объем дисциплины (модуля), час.						Форма текущего контроля успеваемости и, промежуточной аттестации*
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Очная форма обучения								
Тема 4.	Кластеризация.	10			4		6	Т
Тема 5.	Поиск ассоциативных правил.	10			4		6	Т
Тема 6.	Анализ и прогнозирование временных рядов.	10			4		6	Т
Тема 7.	Ансамбли моделей.	10			4		6	Т
Промежуточная аттестация								Зачет
Всего:		72/2			32		40	

*Примечание:*

\* – формы текущего контроля успеваемости: результаты выполнения практической работы на компьютере (КПР), контрольная работа (КР), тестирование (Т).

### 3.1 Содержание дисциплины (модуля)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)
Тема 1	Аналитическая отчетность.	Введение в визуализацию. Визуализаторы общего назначения. OLAP-анализ. Визуализация данных в аналитической платформе Deductor. Компьютерный практикум №1. «Разработка системы аналитической отчетности».
Тема 2	Классификация и регрессия. Статистические методы	Решение задачи классификации в аналитической платформе Deductor. Введение в классификацию и регрессию. Множественная линейная регрессия. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Нелинейные регрессионные модели.

		<p>Основы логистической регрессии. Множественная логистическая регрессия.</p> <p>Оценка эффективности и сравнение моделей. ROC-анализ. Оценка ошибки модели. Издержки ошибочной классификации. Lift- и Profit-кривые.</p> <p>Компьютерный практикум №2. «Скоринговые модели для оценки кредитоспособности заемщиков».</p>
Тема 3	Классификация и регрессия. Машинное обучение	<p>Введение в деревья решений. Алгоритмы построения деревьев решений. Упрощение деревьев решений.</p> <p>Введение в нейронные сети. Принципы построения нейронных сетей. Процесс обучения нейронной сети. Обучение в условиях несбалансированности классов.</p> <p>Компьютерный практикум №3. «Повышение эффективности массовой рассылки».</p>
Тема 4	Кластеризация	<p>Введение в кластеризацию. Меры близости, используемые в алгоритмах кластеризации. Методы кластерного анализа.</p> <p>Иерархическая кластеризация. Алгоритм кластеризации k-means. Сети Кохонена. Карты Кохонена.</p> <p>Решение задачи кластеризации средствами аналитической платформы Deductor.</p> <p>Компьютерный практикум №4. «Сегментация клиентов компании».</p>
Тема 5	Поиск ассоциативных правил	<p>Ассоциативные правила. Алгоритм Apriori. Иерархические ассоциативные правила. Последовательные шаблоны.</p> <p>Компьютерный практикум №5. «Ассоциативные правила в стимулировании розничных продаж».</p>
Тема 6	Анализ и прогнозирование временных рядов	<p>Введение в прогнозирование. Временной ряд и его компоненты. Модели прогнозирования. Исследование наборов данных и выбор метода прогнозирования. Ошибки прогноза. Оценка адекватности выбранного метода прогнозирования.</p> <p>Методы сглаживания и скользящие средние. Декомпозиция временного ряда. Регрессионный анализ временных рядов. Анализ временных рядов средствами аналитической платформы Deductor.</p> <p>Компьютерный практикум №6. «Прогнозирование продаж».</p>
Тема 7	Ансамбли моделей	<p>Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей</p>

#### 4. МАТЕРИАЛЫ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ ОБУЧАЮЩИХСЯ И ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

##### 4.1. Формы и методы текущего контроля успеваемости и промежуточной аттестации

Преподавание дисциплины проводится в форме практических занятий в компьютерных классах (компьютерный практикум).

В компьютерном классе студенты на каждом занятии работают над выполнением практического задания в программе Deductor Academic под руководством преподавателя.

В течение семестра студенты выполняют практическую контрольную работу в программе Deductor Academic.

Каждый вид работы, выполняемый студентом в течение семестра, оценивается рейтинговой оценкой.

4.1.1. В ходе реализации дисциплины «Информационные технологии в прогнозно-аналитической деятельности» используются следующие методы текущего контроля успеваемости обучающихся: тестирование, контрольная работа.

4.1.2. Промежуточная аттестация проводится в форме зачета.

## **4.2. Материалы текущего контроля успеваемости**

### **Примерный вариант вопросов для тестирования к Теме 1:**

1. *Традиционные методы визуализации могут находить следующее применение:*

1. упрощать расчеты в модели
2. компактно описывать закономерности, присущие исходному набору данных
3. представлять пользователю информацию в наглядном виде
4. снижать размерность или сжимать информацию
5. восстановление пробелов в наборе данных

2. *Способы визуального представления могут ...*

1. быть средством оценки качества построенной модели
2. быть иллюстрацией построения модели
3. помочь интерпретировать полученный результат

3. *Какие из перечисленных средств визуализации помогают интерпретировать полученный результат?*

1. дендрограмма
2. дерево решений
3. представление графа нейронной сети

4. *Назовите характеристики одной из основных тенденций в области визуализации:*

1. увеличение размеров структур данных, представляемых визуализацией
2. усложнение структур данных, представляемых визуализацией
3. уменьшение размеров структур данных, представляемых визуализацией

### **Примерный вариант контрольной работы к Теме 1:**

Предлагается решить задачу из области коллекторского скоринга на тему «Оптимизация выбора стратегии взыскания просроченной задолженности».

В файле debtors.txt имеются статистические данные о заемщиках банка, которые



стали должниками, то есть имеют просроченную задолженность. Структура полей файла приведена в следующей таблице.

Имя поля	Описание	Тип
Код	Код заемщика	Целый
Сумма кредита	Сумма кредита в рублях	Вещественный
Срок кредита	Срок кредита в мес. (6, 12, 18)	Целый
Ежемесячный платеж	Размер аннуитетного ежемесячного платежа в рублях	Вещественный
Возраст	Возраст, лет	Целый
Пол	Пол клиента	Строковый
Срок просрочки	Срок просрочки в днях	Целый
Число платежей до выхода на просрочку	Сколько сделал заемщик платежей до того как была впервые зафиксирована просроченная задолженность	Целый
Исполнительный лист	Имеется ли по заемщику исполнительный лист (судебное решение)	0 / 1
Остаток долга (тело)	Сумма в рублях остатка «тела» кредита	Вещественный
Остаток долга (проценты)	Сумма в рублях остатка «процентов» по кредиту	Вещественный
Штрафы	Начисленные штрафы и пенни	Вещественный
Просрочка/Долг	Отношение, рассчитанное на основе трех предыдущих полей по формуле: $(12) / (11 + 10)$	Вещественный
Возобновление платежей	Начал ли заемщик снова оплачивать кредит после работы специалиста коллекторского подразделения: 0 – работа велась, и возобновления платежей не последовало, 1 – работа велась, и последовало возобновление платежей, «пусто» – работа не велась	Целый
Тестовое множество	Признак участия примера в тестовом множестве	Логический

Известна следующая дополнительная информация:

- Ориентировочная стоимость плана мероприятий по взысканию просроченной задолженности с 1 заемщика оценивается в 1 000 рублей ((зарплата персоналу, оплата звонков, выездов и т.д.)
- Средняя ожидаемая сумма погашения от 1 должника 12 000 рублей.

#### Задание:

1. Построить модели, позволяющие классифицировать должников на 2 класса в зависимости от исхода мероприятий, проводимого специалистом коллекторского подразделения банка (логистическая регрессия, дерево решений, многослойный пер-

септрон). Разделение на обучающее и тестовое множество уже проведено – используйте поле Тестовое множество. За положительный исход примите Возобновление платежей = 1.

2. Построить Lift-кривые для моделей (там, где это возможно) и сравнить модели.
3. Используя дополнительную информацию, выбрать лучшую модель с двух позиций:
  - a. Максимизация ожидаемых поступлений от должников;
  - b. Минимизация рабочего времени, которое будет потрачено на мероприятия по взысканию.
4. «Прогоните» через лучшую модель тех должников, к которым еще не применялись мероприятия по взысканию и получите список клиентов, с которыми нужно работать, ранжированный по убыванию вероятности положительного исхода.
5. Все важные визуализаторы вывести на панель отчетов. Свои комментарии впишите к любому узлу-обработчику.
6. Ответьте на вопрос: какой из алгоритмов оказался наименее чувствителен к несбалансированности классов? Комментарии впишите к любому узлу-обработчику.

### **Примерный вариант вопросов для тестирования к Теме 2:**

5. Проставьте соответствие между классами задач Data Mining и их определениями:

Классификация	выявление закономерностей между связанными событиями
Регрессия	установление зависимости непрерывной выходной переменной от входных переменных
Кластеризация	группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов
Ассоциация	установление зависимости дискретной выходной переменной от входных переменных

6. К какой форме данных относится информация на веб-странице в Интернете?

1. структурированные
2. слабоструктурированные
3. неструктурированные

7. С какими формами данных, как правило, работают методы анализа данных?

1. структурированные
2. неструктурированные и слабоструктурированные
3. слабоструктурированные
4. неструктурированные
5. структурированные и слабоструктурированные

8. Проставьте соответствие между пропущенными словами и текстом:

[...] данные отражают отдельные факты предметной области.	Слабоструктурированные
---	------------------------

[...] данные имеют некоторые правила и форматы их задания, но в самом общем виде.	Неструктурированные
[...] данные являются произвольными по форме.	Структурированные

9. Можно ли статистический пакет отнести к классу аналитических платформ?

1. Да
2. Нет

10. Кто является основателем направления Knowledge Discovery in Databases и Data Mining?

1. Л. Заде
2. Б. Инмон
3. Г. Пятецкий-Шапиро
4. Б. Коско

11. Каковы особенности данных, накапливаемых в компаниях? (Выберите три варианта)

1. Данные редко накапливаются специально для решения задач анализа
2. Данные всегда представлены в структурированной форме
3. Как правило, данные содержат ошибки, аномалии и пропуски
4. Нередко имеют большой объем
5. Почти всегда носят неполный, фрагментарный характер

12. Могут ли числовые данные быть дискретного вида при решении задачи анализа?

1. Нет
2. Да

13. Выберите неверный вариант:

1. Эксперт выдвигает гипотезы и строит модели для проверки достоверности гипотез
2. Аналитик – это специалист в области анализа и моделирования
3. Эксперт – это специалист предметной области, профессионал, который за годы обучения и практической деятельности научился эффективно решать задачи, относящиеся к конкретной предметной области
4. Эксперт является связующим звеном между специалистами разных уровней и областей

14. Проставьте соответствие между пропущенными словами и текстом:

При [...] подходе модель не «подстраивается» под действительность, а исследователь пытается подобрать модель	информационном
--	----------------

таким образом, чтобы она адекватно отражала реальность.	
В [...] подходе отправной точкой являются данные, характеризующие исследуемый объект, и модель «подстраивается» под действительность.	аналитическом

15. Укажите основные виды моделей:

1. эмпирические
2. теоретические
3. смешанные
4. экспериментальные
5. статистические

16. Как переводится *Knowledge Discovery in Databases (KDD)*?

1. извлечение знаний из баз данных
2. тиражирование знаний
3. «раскопка» данных

17. Отметьте неверное рассуждение:

1. Под моделированием понимается процесс исследования реальной системы, включающий построение модели, изучение ее свойств и перенос полученных сведений на моделируемую систему
2. Построение моделей является единственным способом изучения окружающего мира
3. Модель – это объект, который имеет сходство в некоторых отношениях с прототипом и служит средством описания и объяснения, прогнозирования поведения прототипа
4. Модель – это объект или описание объекта, системы для замещения одной системы другой системой для лучшего изучения оригинала или воспроизведения каких-либо его свойств.

18. Какие существуют основные способы решения аналитических задач?

1. Построение и использование моделей
2. Верны оба варианта
3. Извлечение и визуализация данных

19. Какая самая распространенная модель хранения структурированных данных?

1. текст
2. граф
3. дерево
4. матрица

таблица

**Примерный вариант вопросов для тестирования к Теме 3:**

20. *Классификация — это ...*
  1. отнесение объектов к одной из заранее неизвестных групп
  2. процесс формирования групп и отнесения объектов к одному из них.
  3. отнесение объектов к одному из заранее известных классов
  
21. *Задачи классификации решаются следующими алгоритмами:  
(Ответ считается верным, если отмечены все правильные варианты ответов.)*
  1. алгоритмом Apriori
  2. нейронные сети
  3. линейной регрессии
  
22. *Деревья решений относятся к группе (-ам) ...*
  1. кибернетических методов
  2. методов кросс-табуляции
  3. статистических методов
  4. логических методов
  
23. *Внутренний узел дерева решений называют также ...*
  1. узлом проверки
  2. конечным узлом
  3. вершиной
  4. листом
  
24. *Конечный узел дерева решений называют также ...*
  1. узлом проверки
  2. узлом решения
  3. листом
  
25. *При помощи метода деревьев решений решаются задачи ...*
  1. классификации и кластеризации
  2. классификации и прогнозирования
  3. кластеризации, классификации и прогнозирования
  4. кластеризации и прогнозирования
  
26. *Процесс отсечения ветвей или замена некоторых ветвей поддеревом ...*
  1. происходит сверху вниз, т.е. является нисходящим
  2. происходит снизу вверх, т.е. является восходящим
  3. может быть как нисходящим, так и восходящим

27. *Алгоритм конструирования дерева решений ...*
  1. не требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых
  2. требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых
  3. на вход алгоритма можно подавать все существующие атрибуты, алгоритм сам выберет наиболее значимые среди них, и только они будут использованы для построения дерева
  
28. *Если зависимая переменная принимает непрерывные значения, то дерево решений решает задачу:*
  1. классификации и численного прогнозирования
  2. численного прогнозирования
  3. классификации
  
29. *Если зависимая переменная принимает дискретные значения, при помощи метода дерева решений решается задача:*
  1. классификации и численного прогнозирования
  2. классификации
  3. численного прогнозирования
  
30. *Метод, который делает заключения относительно данной ситуации по результатам поиска аналогий, хранящихся в базе прецедентов относится к категории ...*
  1. обучение с учителем"
  2. самообучающейся системы
  3. "обучение без учителя"
  
31. *Явление переобучения характеризуется ...*
  1. чрезмерно точным соответствием нейронной сети конкретному набору обучающих примеров, при котором сеть теряет способность к обобщению
  2. возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры нейронной сети
  3. возникновением, в случае слишком долгого обучения, слишком большого числа обучающих примеров или слишком сложной структуры нейронной сети
  
32. *Многослойный персептрон – это сеть ...*
  1. обратного распространения сигнала (с обратными связями)
  2. в которой входной сигнал преобразуется в выходной, проходя последовательно через несколько слоев
  3. прямого распространения сигнала (без обратных связей)
  
33. *В многослойном персептроне ...*
  1. должен быть хотя бы один скрытый слой

2. присутствие нескольких скрытых слоев оправдано лишь в случае использования нелинейных функций активации
  3. может быть какое угодно количество скрытых слоев, они также могут вообще отсутствовать
34. *Заполните пропуски в формулировке: "... – выходная связь нейрона, с которой сигнал (возбуждения или торможения) поступает на... следующих нейронов"*
1. синапс, аксоны
  2. аксон, точку ветвления
  3. аксон, синапсы
35. *Однонаправленные входные связи, соединенные с выходами других нейронов – это ...*
1. синапсы
  2. слои сети
  3. аксоны
36. *Слой нейронной сети – это ...*
1. выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов
  2. один или несколько нейронов, на входы которых подается один и тот же общий сигнал
  3. однонаправленные входные связи, соединенные с выходами других нейронов
37. *Ошибкой обучения нейронной сети называется ...*
1. переобучение нейронной сети
  2. разность между желаемым и полученным на выходе сигналами
  3. целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети
38. *Целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети – это ...*
1. функция ошибок
  2. функция переобучения
  3. ошибка обучения

#### **Примерный вариант вопросов для тестирования к Теме 4:**

39. *Что такое мощность кластера?*
1. яркость цветовой раскраски кластера
  2. длина окружности кластера
  3. число объектов, попавших в кластер
  4. форма кластера
  5. радиус кластера
40. *Радиус обучения при подборе весов сети Кохонена с каждой итерацией:*
1. увеличивается
  2. уменьшается
  3. остается неизменным

41.  $D = \sum_i |x_i - y_i|$       Формула соответствует:
1. функции отличия
  2. расстоянию Манхэттена
  3. расстоянию Уорда
  4. расстоянию Евклида
42. Какой нейрон в алгоритме Кохонена объявляется «победителем»?
1. вектор весов которого имеет наименьшее расстояние до вектора признаков объекта
  2. вектор весов которого равноудален от суммы квадратов всех входных нейронов
  3. вектор весов которого имеет минимальное число связей с входными нейронами
43. При построении карты Кохонена поле Пол – выходное. В кластер попали 56 женщин и 44 мужчины. Новый объект, попавший в этот кластер, будет отнесен к:
1. женщине
  2. неизвестно
  3. мужчине
44. Кластеризация, как правило, является ... этапом при анализе данных и поиске закономерностей
1. конечным
  2. новым
  3. главным
  4. подготовительным
45. Отметьте неверные суждения
1. При выполнении кластеризации важно, сколько в результате должно быть построено кластеров.
  2. Из-за сложности разработки существует небольшое число алгоритмов кластеризации.
  3. Кластеризация сама по себе представляет собой ценный результат.
  4. Кластерная модель должна описывать как сами кластеры, так и принадлежность каждого объекта к одному из них.
  5. Результат кластеризации практически не зависит от выбранной меры близости между объектами.
  6. Не существует единственного «правильного» решения задачи кластеризации.
46. Можно ли при помощи кластеризации решать задачи прогнозирования?
1. нет
  2. да
47. Такие алгоритмы кластеризации, как *k-means* и сети Кохонена ... приспособлены для работы с наборами данных, состоящих преимущественно из категориальных признаков
1. хорошо
  2. плохо
48. Какой способ инициализации весов сети Кохонена применяется по умолчанию?
1. из собственных векторов
  2. из обучающего множества
  3. случайными значениями



49. *Правило формирования окрестности (несколько нейронов, которые окружают нейрон-победитель):*
1. размер окрестности в процессе обучения не изменяется
  2. сначала к окрестности принадлежит малое число нейронов, далее ее размер постепенно увеличивается
  3. сначала к окрестности принадлежит большое число нейронов, далее ее размер постепенно уменьшается
50. *Закончите фразу: "В процессе обучения сетей Кохонена на входы подаются данные, сеть при этом подстраивается..."*
1. не под закономерности во входных данных, а под коэффициенты весов
  2. не под эталонное значение выхода, а под закономерности во входных данных
  3. не под закономерности во входных данных, а под эталонное значение выхода
51. *Спорный объект кластеризации — это объект, который по мере сходства ...*
1. может быть отнесен к нескольким кластерам
  2. не может быть отнесен ни к одному кластеру
  3. может быть отнесен более чем к двум кластерам
52. *Иерархические дивизимные методы характеризуются ...*
1. сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
  2. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
  3. делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп
53. *Назовите характеристики кластерного анализа:*
1. накладывает ограничения на представление исследуемых объектов
  2. не накладывает ограничения на представление исследуемых объектов
  3. требует априорных предположений о наборе данных
  4. не требует априорных предположений о наборе данных
  5. позволяет анализировать показатели различных типов данных.
54. *Какой метод рекомендуется использовать при небольших объемах выборки?*
1. метод ближнего соседа
  2. все ответы неверны
  3. метод k-средних
- вся группа иерархических методов

### **Примерный вариант вопросов для тестирования к Теме 5:**

55. *В ходе решения какой из перечисленных задач устанавливаются закономерности между связанными событиями в наборе данных?*
1. задачи поиска последовательных ассоциативных правил
  2. задачи поиска ассоциативных правил
  3. задачи анализа отклонений
56. *Из общего количества 1000 покупок в магазине было приобретено 300 мобильных телефонов, а 100 человек из купивших телефон приобрели и чехол к нему. Поддержка правила Телефон > Чехол равна*

%

57. *Задача ассоциации впервые возникла*
1. при анализе веб - страниц
  2. в торговле при анализе рыночной корзины
  3. при исследовании действия побочных эффектов лекарств
58. *Часто встречающиеся множество или популярный предметный набор это:*
1. предметный набор с достоверностью, больше либо равной заданного порога
  2. предметный набор с поддержкой, больше либо равной заданного порога
  3. предметный набор с достоверностью, меньше либо равной заданного порога
  4. предметный набор с поддержкой, меньше либо равной заданного порога
59. *Один из первых популярных алгоритмов генерации ассоциативных правил это:*
1. back propagation
  2. FPG
  3. SOM
  4. a priori
  5. a posteriori
60. *В ассоциативном правиле  $A \Rightarrow B$  вероятность того, что из наличия в транзакции товара A следует наличие в ней товара B показывает значение*
1. левереджа
  2. улучшения
  3. лифта
  4. поддержки
  5. достоверности
61. *Отметьте неверные суждения:*
1.  $S(\text{не } A) = 1 - S(A)$ , где  $S$  – поддержка набора A.
  2. лифт ассоциативного правила  $A \Rightarrow B$  равен лифту правила  $B \Rightarrow A$
  3. достоверность позволяет оценить полезность правила
  4. иерархические ассоциативные правила позволяют решить проблему, когда ассортимент товаров очень велик и ассоциации с высокой поддержкой для многих товаров отсутствуют
  5. лифт ассоциативного правила показывает, какой процент транзакций поддерживает данное правило
62. *Сколько входных полей должно быть (т.е. подаваться на вход алгоритма) в обработчике Deductor Ассоциативные правила?*
1. 2
  2. 1
  3. 3
  4. (количество предметов в транзакциях) плюс 1

#### **Примерный вариант вопросов для тестирования к Теме 6:**

63. *Параметрами прогнозирования являются:*
1. интервал прогнозирования
  2. тренд
  3. горизонт прогнозирования
  4. период прогнозирования

64. *Период прогнозирования – это ...*
1. характеристика временного ряда
  2. параметр прогнозирования
  3. составляющая временного ряда
65. *Какой из параметров является основной единицей времени, на которую делается прогноз?*
1. горизонт прогнозирования
  2. интервал прогнозирования
  3. период прогнозирования
66. *Продолжите фразу: "Прогнозирование будет иметь смысл, если горизонт прогнозирования ..."*
1. не больше, чем время, которое необходимо для реализации решения, принятого на основе прогноза"
  2. не меньше, чем время, которое необходимо для реализации решения, принятого на основе прогноза"
  3. оба ответа верны
67. *Какой из параметров является частотой, с которой делается новый прогноз?*
1. период прогнозирования
  2. горизонт прогнозирования
  3. интервал прогнозирования
68. *В чем состоит основное сходство задач прогнозирования и классификации?*
1. сходство заключается в том, что при решении обеих задач предсказываются числовые значения зависимой переменной
  2. при решении обеих задач используется двухэтапный процесс построения модели на основе обучающего набора и ее использования для предсказания неизвестных значений зависимой переменной
  3. оба ответа верны
69. *В чем состоит основное отличие задач прогнозирования и классификации?*
1. отличие задач классификации и прогнозирования состоит в том, что в первой задаче предсказывается класс зависимой переменной, а во второй - числовые значения зависимой переменной
  2. отличие заключается в этапах процесса решения задач

#### **Примерный вариант вопросов для тестирования к Теме 7:**

70. *Назовите характеристики одной из основных тенденций в области визуализации:*
4. увеличение размеров структур данных, представляемых визуализацией
  5. усложнение структур данных, представляемых визуализацией
  6. уменьшение размеров структур данных, представляемых визуализацией
71. *На какие две группы подразделяются методы Data Mining по принципу работы с исходными обучающими данными?*
1. статистические методы
  2. непосредственное использование данных или сохранение данных
  3. кибернетические методы

4. выявление и использование формализованных закономерностей
72. *Какие из перечисленных ниже пунктов являются названиями стадий Data Mining?*
  1. индукция правил
  2. свободный поиск
  3. анализ исключений
  4. прогностическое моделирование
73. *Любые, неизвестные ранее сведения о каком-либо событии, сущности, процессе и т.п., являющиеся объектом некоторых операций, для которых существует содержательная интерпретация, являются...*
  1. знаниями
  2. информацией
  3. данными
74. *У основания так называемой информационной пирамиды находится категория ...*
  1. информация
  2. знания
  3. данные
75. *Задачи Data Mining, в зависимости от используемых моделей подразделяются на:*
  1. обучение без учителя
  2. прогнозирующие
  3. дескриптивные
  4. обучение с учителем
76. *Продолжите фразу: "Кластеризация и классификация относятся к..."*
  1. стратегии обучения с учителем"
  2. стратегии обучения без учителя"
  3. к двум разным стратегиям: обучения без учителя и обучения с учителем
77. *В результате использования одних и тех же данных и различных методов...*
  1. должна появляться только одинаковая информация
  2. может появляться разная информация, это зависит от выбранных методов обработки данных
  3. должна появляться разная информация
78. *В ходе решения какой из перечисленных задач устанавливаются закономерности между связанными событиями в наборе данных?*
  4. задачи поиска последовательных ассоциативных правил
  5. задачи поиска ассоциативных правил

задачи анализа отклонений

### 4.3. Оценочные средства для промежуточной аттестации

**4.3.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы. Показатели и критерии оценивания компетенций с учетом этапа их формирования**

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ОПК-7	Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учётом основных требований информационной безопасности	ОПК-7.3	Способность решить стандартные задачи профессиональной деятельности на основе информационной культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
ОПК-7.3 Способность решить стандартные задачи профессиональной деятельности на основе информационной культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	Способен оценить необходимость и возможность применения интеллектуального анализа данных при решении конкретных бизнес-задач; Способен осуществлять выбор методов анализа, подготовку исходных данных, визуализацию и интерпретацию полученных результатов; Способен решать задачи ассоциации, кластеризации, классификации, регрессии, задачи анализа и прогнозирования с помощью временных рядов, строить аналитическую отчетность.	Проведена оценка возможности применения интеллектуального анализа данных при решении конкретных бизнес-задач; Сделан выбор методов анализа, подготовлены исходные данные, проведена интерпретация полученных результатов; Решена задача анализа и прогнозирования с помощью временных рядов, подготовлена аналитическая отчетность.

#### 4.3.2. Типовые оценочные средства

##### Вопросы для подготовки к зачету

1. Аналитический и информационный подход к моделированию.
2. Основные этапы построения моделей.
3. Формы представления, типы и виды данных.
4. Классификация задач Data Mining.

5. Виды источников данных.
6. Хранилища данных.
7. Многомерная модель данных.
8. Оценка качества данных.
9. Предобработка данных.
10. Трансформация данных.
11. Цели и задачи визуализации данных в аналитических технологиях.
12. Виды визуализаторов данных.
13. OLAP-анализ.
14. Задача поиска ассоциаций.
15. Значимость и полезность ассоциативных правил.
16. Алгоритм apriori.
17. Задача кластеризации.
18. Алгоритмы решения задачи кластеризации.
19. Алгоритм k-means.
20. Сети и карты Кохонена.
21. Задача классификации.
22. Статистические методы решения задачи классификации.
23. Деревья решений.
24. Нейронные сети.
25. Эффект переобучения.
26. Оценка качества классификационных моделей.
27. Модели временных рядов.
28. Методы сглаживания временных рядов.
29. Оценка адекватности выбранного метода прогнозирования.
30. Ансамбли моделей. Бэггинг.
31. Ансамбли моделей. Бустинг.

### **Примеры зачетных задач**

#### *Вариант 1.*

На основе данных хранилища данных Банк получить следующие отчеты и вынести их на Панель Отчетов.

1. Количество одобренных/отклоненных заявок по каждому филиалу и отделению на основе всех имеющихся данных.

2. 10 «наиболее злостных» неплательщиков по состоянию на 10.11.2009
3. Среднюю интенсивность поступления заявок по дням недели каждого из отделений банка на основании всех имеющихся данных. Отчет дополните соответствующим графиком.

*Вариант 2.*

Импортировать текстовый файл Loans\_demo.txt.

С помощью обработки Логистическая регрессия построить модель, которая могла бы предсказывать наличие в будущем у заемщика просрочек по обслуживанию долга.

1. Оценить качество модели.
2. Чему равен оптимальный порог отсечения?
3. Как изменяется кредитный рейтинг заемщика с каждым дополнительным иждивенцем?
4. К какому классу будут отнесен заемщик  
Мужчина, 43-х лет, состоит в браке, 3 иждивенца, 20500 доход, опыт работы 20, срок проживания 10, недвижимость 0, месячный платеж 6725.

*Вариант 3.*

Импортировать текстовый файл responses.txt.

С помощью обработок «Дерево решений» и «Нейронная» сеть построить модели позволяющие предсказать отклик клиента на рассылку.

1. Сравнить качество моделей.
2. Какую модель для отбора респондентов из рассмотренных вы выберете, если известно, что бюджет рассылки составляет 95 000 руб.?
3. Скольким клиентам пришлось бы отправить письма, если бы мы проводили рассылку случайно? Какую экономию дает наша модель?

*Вариант 4.*

Импортировать текстовый файл mobile.txt.

1. Провести сегментацию абонентов оператора мобильной связи.
2. Построить профили абонентов путем выявления их схожего поведения в плане частоты, длительности и времени звонков, а так же ежемесячных расходов;
3. Провести оценку наиболее и наименее доходных сегментов.

*Вариант 5.*

Исходные данные представлены в файле Чеки.

Стоят следующие задачи:

1. предсказать то, какие товары покупатели могут выбрать в зависимости от того, что уже есть в их корзинах;
2. выявить наиболее популярные товарные наборы, состоящие из более чем 1 предмета;
3. предложить рекламные акции типа «каждому купившему А и В товар С в подарок»;

*Вариант 6.*

На основе данных хранилища данных Materials

1. построить модель, позволяющую автоматизировать ежемесячный расчет товарного прогноза на следующие три периода. Типовой сценарий построить на примере товарной группы «Грунтовка».
2. Оценить качество предложенной модели.

### Шкала оценивания зачета

	выставляется студенту, если он представил более чем 90% правильных ответов решения предложенных на зачете задач, может проанализировать полученные результаты, объяснить ход решения задачи, отстаивать свою точку зрения;	40 баллов
	выставляется студенту, если он представил более чем 75% правильных ответов решения предложенных на зачете задач, может проанализировать полученные результаты, объяснить ход	30 баллов



	решения задачи, отстаивать свою точку зрения	
	выставляется студенту, если он представил более чем 50% правильных ответов решения предложенных на зачете задач, может проанализировать полученные результаты, но при расчетах допускает арифметические и методические ошибки	20 баллов
	выставляется студенту, если он представил менее чем 50% правильных ответов решения предложенных на зачете задач, при расчетах допускает грубые арифметические и методические ошибки	0 баллов

### Шкала оценивания по дисциплине

Наименование темы (раздела)	Показатели оценивания	Критерии оценивания См. п.4.4.	Оценка (баллы)
Аналитическая отчетность.	Тестирование Контрольная работа	Ответы на все вопросы Правильность выполненного задания и срок выполнения	<b>12</b>
Классификация и регрессия. Статистические методы	Тестирование	Ответы на все вопросы	<b>8</b>
Классификация и регрессия. Машинное обучение	Тестирование	Ответы на все вопросы	<b>8</b>
Кластеризация	Тестирование	Ответы на все вопросы	<b>8</b>

Поиск ассоциативных правил	Тестирование	Ответы на все вопросы	<b>8</b>
Анализ и прогнозирование временных рядов	Тестирование	Ответы на все вопросы	<b>8</b>
Ансамбли моделей	Тестирование	Ответы на все вопросы	<b>8</b>
Все темы		зачет	<b>40</b>

Шкала перевода оценки знаний по 100-балльной системе в эквивалент пятибалльной системы приведена в таблице:

Шкала перевода оценок

Баллы по 100-балльной системе	Пятибалльная система оценки
85-100 баллов	Отлично (зачет)
70-84 баллов	Хорошо (зачет)
50-69 баллов	Удовлетворительно (зачет)
49 балл и ниже	Неудовлетворительно (незачет)

#### 4.4. Методические материалы

##### Шкала оценивания тестирования

№ п/п	Критерии оценивания	Баллы
1	Выполнено не менее 90% задания	8
2	Выполнено не менее 80% задания	6
4	Выполнено не менее 60% задания	4
5	Выполнено не менее 50% задания	2
6	Выполнено менее 50% задания	0

### Критерии оценки выполнения контрольной работы

Критерии оценки: - соответствие предполагаемым ответам; - правильное использование алгоритма выполнения действий (методики, технологии и т.д.); - логика рассуждений; - неординарность подхода к решению.		макс 4 баллов
	выставляется студенту, если он владеет категориальным аппаратом, может привести классификацию факторов явления, решить поставленную задачу и проанализировать полученные результаты, объяснить причины отклонений от желаемого результата, отстаивать свою точку зрения, приводя факты;	(4) баллов
	выставляется студенту, если он владеет категориальным аппаратом, может привести классификацию факторов явления, решить поставленную задачу и проанализировать полученные результаты;	(3) баллов
	выставляется студенту, если он владеет категориальным аппаратом, может привести формулы расчета, рассчитать задание;	(2) баллов
«2»	выставляется студенту, если он не владеет перечисленными навыками	(0 баллов).

## 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

### Методические рекомендации по выполнению тестов:

- Внимательно изучите структуру теста, оцените объем времени, выделяемого на данный тест, посмотрите, какого типа задания в нем содержатся.
- Начинайте отвечать на те вопросы, в правильности решения которых нет сомнений, пока не останавливаясь на тех, которые могут вызвать долгие раздумья. Это позволит успокоиться и сосредоточиться на выполнении более трудных вопросов.
- Всегда внимательно читайте задания до конца, не пытаясь понять условия «по первым словам» или выполнив подобные задания в предыдущих тестированиях. Такая спешка нередко приводит к досадным ошибкам в самых легких вопросах.
- Если Вы не знаете ответа на вопрос или не уверены в правильности, следует пропустить его и отметить, чтобы потом к нему вернуться.
- Психологи также советуют думать только о текущем задании. Как правило, задания в тестах не связаны друг с другом непосредственно, поэтому необходимо концентрироваться на данном вопросе и находить решения, подходящие именно к нему. Кроме того, выполнение этой рекомендации даст еще один психологический эффект – позволит забыть о неудаче в ответе на предыдущий вопрос, если таковая имела место.
- Многие задания можно быстрее решить, если не искать сразу правильный вариант ответа, а последовательно исключать те, которые явно не подходят. Метод исключения позволяет в итоге сконцентрировать внимание на одном-двух вероятных вариантах.
- Рассчитывать выполнение заданий нужно всегда так, чтобы осталось время на проверку и доработку (примерно 1/3-1/4 запланированного времени). Тогда вероятность опуск сводится к нулю и имеется время, чтобы набрать максимум баллов на легких заданиях и сосредоточиться на решении более трудных, которые вначале пришлось пропустить.
- Процесс угадывания правильных ответов желательно свести к минимуму, так как это чревато тем, что студент забудет о главном: умении использовать имеющиеся накопленные в учебном процессе знания, и будет надеяться на удачу. Если уверенности в правильности ответа нет, но интуитивно появляется предпочтение, то психологи рекомендуют доверять интуиции, которая считается проявлением глубинных знаний и опыта, находящихся на уровне подсознания.

При подготовке к тесту не следует просто заучивать, необходимо понять логику изложенного материала. Этому немало способствует составление развернутого плана, таблиц, схем, внимательное изучение исторических карт. Большую помощь оказывают опубликованные сборники тестов, Интернет-тренажеры, позволяющие, во-первых, закрепить знания, во-вторых, приобрести соответствующие психологические навыки саморегуляции и самоконтроля. Именно такие навыки не только повышают эффективность подготовки, позволяют более успешно вести себя во время экзамена, но и вообще способствуют развитию навыков мыслительной работы.

## **6. УЧЕБНАЯ ЛИТЕРАТУРА И РЕСУРСЫ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», ВКЛЮЧАЯ ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ**

### **6.1 Основная литература**

1. Александровская Ю.П. [и др.]. Информационные технологии в экономике и управлении [Электронный ресурс]: учебное пособие/— Электрон. текстовые данные.— Казань: Казанский национальный исследовательский технологический университет, 2016.— 112 с.— Режим доступа: <http://www.iprbookshop.ru/61853.html>.— ЭБС «IPRbooks»
2. Миркин Б.Г. Введение в анализ данных. — М.: Издательство Юрайт, 2017. — 174 с. <https://www.biblio-online.ru/book/46A41F93-BC46-401C-A30E-27C0FB60B9DE>
3. Уткин В.Б. Информационные системы и технологии в экономике [Электронный ресурс]: учебник для вузов/ В.Б. Уткин, К.В. Балдин— Электрон. текстовые данные.— М.: ЮНИТИ-ДАНА, 2016.— 336 с.— Режим доступа: <http://www.iprbookshop.ru/7040.html>.— ЭБС «IPRbooks»

### **6.2 Дополнительная литература**

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. — СПб.: Питер, 2016. — 704 с.
2. Подкорытова О.А., Соколов М.В. Анализ временных рядов. — М.: Издательство Юрайт, 2017. — 266 с. <https://www.biblio-online.ru/book/7132122F-D176-4118-AD03-D43A9FA2FF86>

### **6.3 Интернет-ресурсы**

1. <http://www.intuit.ru/studies/curriculums/6/courses/6/info> — Национальный Открытый Университет «ИНТУИТ».

## **7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ**

Лекционные занятия проводятся в учебных аудиториях, оснащенных рабочим местом преподавателя (стол, стул, кафедра), рабочими местами студентов (столы, стулья) по количеству студентов, доской меловой или белой для написания маркерами или флипчартом для бумаги большого формата, маркерами (красный, черный, зеленый, синий), губкой для досок, оборудованием для показа презентаций и слайдов (компьютер, проектор, экран).

Практические занятия проводятся в компьютерном классе, оснащенном персональными компьютерами, объединенными в сеть с выходом в Интернет, а также оборудованием для показа презентаций и слайдов (компьютер, проектор, экран).

Используется следующее программное обеспечение:

Microsoft Windows 10 LTSC 1607

Количество 2607

Правообладатель Microsoft Corporation

Дата покупки / продления 06.12.2016

Контракт 59/07-16/0373100037616000052-0008121-03

Продавец ООО «ЛАНИТ-Интеграция»

Покупатель РАНХиГС

Дата окончания 31.12.2017

Срок подписки 1 год / 3 года

Microsoft Office Professional 2016

Количество 2607

Правообладатель Microsoft Corporation

Дата покупки / продления 06.12.2016

Контракт 59/07-16/0373100037616000052-0008121-03

Продавец ООО «ЛАНИТ-Интеграция»

Покупатель РАНХиГС

Дата окончания 31.12.2017

Срок подписки 1 год / 3 года

Acrobat Professional AcademicEdition License Russian

Multiple Platforms (Adobe, 65258631AE01A00)

Количество 50

Правообладатель Adobe

Дата покупки / продления 03.04.2017

Контракт #15/08-17

Продавец SoftLine

Покупатель РАНХиГС

Дата окончания 03.04.2018

Для выполнения практических заданий по дисциплине используется свободно распространяемое программное обеспечение Deductor Academic.