

Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)

Департамент логистики и маркетинга
Факультета экономики и бизнеса

И.А. МЕРКУЛИНА

МЕТОДЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ В ЛОГИСТИКЕ

Рабочая программа дисциплины
для студентов, обучающихся по направлению подготовки
38.04.02 «Менеджмент»,
направленность программы магистратуры
«Логистика: финансовые и цифровые технологии»

Москва 2022

Федеральное государственное образовательное бюджетное учреждение
высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)

Департамент логистики и маркетинга
Факультета экономики и бизнеса

УТВЕРЖДАЮ
Проректор по учебной и
методической работе

_____ Е.А. Каменева

26 декабря 2022 г.

И.А. МЕРКУЛИНА

МЕТОДЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ В ЛОГИСТИКЕ

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

38.04.02 «Менеджмент»,

направленность программы магистратуры

«Логистика: финансовые и цифровые технологии»

Рекомендовано Ученым советом Факультета экономики и бизнеса

(протокол № 25 от 21.12.2022)

Одобрено Советом учебно-научного Департамента логистики и маркетинга

(протокол № 3 от 19.12.2022)

Москва 2022

СОДЕРЖАНИЕ

1. Наименование дисциплины	4
2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения, соотнесенных с планируемыми результатами обучения по дисциплине	4
3. Место дисциплины в структуре образовательной программы	5
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	6
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	6
5.1. Содержание дисциплины	6
5.2. Учебно-тематический план	7
5.3. Содержание семинаров, практических занятий	9
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	10
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	10
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	12
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	19
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	29
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	30
10. Методические указания для обучающихся по освоению дисциплины	29
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	32
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	38

1. Наименование дисциплины

«Методы анализа больших данных в логистике».

2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения, соотнесенных с планируемыми результатами обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-2	Способность применять современные методы и техники сбора, обработки и анализа данных, а также определения и прогнозирования основных социально-экономических показателей объектов управления	1. Разрабатывает методы, техники и инструментарий для анализа и прогнозирования тенденций и социально-экономических показателей	<ul style="list-style-type: none"> • Знать: особенности анализа и прогнозирования тенденций и социально-экономических показателей и технико-методический инструментарий их разработки • Уметь: применять методы, техники и инструментарий для анализа и прогнозирования тенденций и социально-экономических показателей
		2. Использует инструменты диагностики изменения состояния объектов управления на ранних стадиях в целях прогнозирования результатов их деятельности и предотвращения негативных последствий	<ul style="list-style-type: none"> • Знать: возможные варианты состояния объектов управления на ранних стадиях, методы прогнозирования результатов их деятельности и способы предотвращения негативных последствий • Уметь: применять инструменты диагностики изменения состояния объектов управления на ранних стадиях в целях прогнозирования результатов их деятельности и предотвращения негативных последствий
		3. Владеет способностью анализировать проблемы финансово-экономического состояния организаций и прогнозировать их последствия.	<ul style="list-style-type: none"> • Знать: методы проблемно-ориентированного анализа финансово-экономического состояния логистических организаций и механизмы оценки последствий диагностированного состояния • Уметь: применять анализ финансово-экономического состояния логистических организаций в целях повышения эффективности их функционирования

		4. Применяет интеллектуальные информационные технологии для повышения эффективности управления знаниями.	<ul style="list-style-type: none"> • Знать: виды и типологию интеллектуальных информационных технологий, применяемых для повышения эффективности управления знаниями в логистике • Уметь: проводить оценку эффективности управления знаниями в логистике с применением интеллектуальных информационных технологий
ПК-3	Способность разрабатывать экономические, финансовые и математические модели бизнес-процессов в логистике	1. Владеет различными методиками моделирования бизнес-процессов логистических компаний	<ul style="list-style-type: none"> • Знать: способы и этапы моделирования бизнес-процессов, инициированных логистическими компаниями • Уметь: применять основные методы моделирования бизнес-процессов в логистических компаниях
		2. Применяет методы разработки экономических, финансовых и математических моделей для создания и развития логистического бизнеса	<ul style="list-style-type: none"> • Знать: возможности применения экономических, финансовых и математических моделей в логистическом бизнесе. • Уметь: применять методы разработки экономических, финансовых и математических моделей для создания и развития логистического бизнеса
		3. Разрабатывает предложения по повышению эффективности бизнес-процессов в логистике с применением цифровых технологий	<ul style="list-style-type: none"> • Знать: цифровые технологии, применение которых направлено на повышение эффективности логистических бизнес-процессов • Уметь: управлять разработкой эффективных экономических решений по рациональной организации бизнес-процессов в логистике на основе цифровых технологий

3. Место дисциплины в структуре образовательной программы

Дисциплина «Методы анализа больших данных в логистике» относится к модулю дисциплин по выбору, углубляющих освоение программы магистратуры «Логистика: финансовые и цифровые технологии» по направлению подготовки 38.04.02 «Менеджмент».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Таблица 1

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Модуль 6 (в часах)
Общая трудоемкость дисциплины	3 з.е./108	108
Контактная работа - Аудиторные занятия	30	30
<i>Лекции</i>	10	10
<i>Семинары, практические занятия</i>	20	20
Самостоятельная работа	78	78
Вид текущего контроля	<i>Домашнее творческое задание</i>	<i>Домашнее творческое задание</i>
Вид промежуточной аттестации	<i>Зачёт</i>	<i>Зачёт</i>

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Методы применения Data Mining и Knowledge discovery in databases в логистике

Теоретические основы Data Mining. Теория баз данных, прикладной статистики и машинного обучения. Стратегии извлечения данных. Технологии Data Mining: статистический анализ, деревья решений, нейросети, методы поиска логических правил, методы визуализации найденных закономерностей и др. Теоретические основы knowledge discovery in databases «обнаружение знаний в базах данных».

Тема 2. Методы интеллектуального анализа больших данных в логистике

Задачи интеллектуального анализа данных в логистике. Поиск зависимостей (ассоциации, последовательности). Выделение классов (классификация, кластеризация) в логистике. Описание концепций (обобщение, распознавание, сравнение). Обнаружение отклонений (обнаружение аномалий, выбросов; анализ тенденций). Визуализация данных.

Тема 3. Классы программных средств интеллектуального анализа больших данных в логистике

Категории систем интеллектуального анализа данных в логистике. Предметно-ориентированные аналитические системы. Статистические пакеты. Системы визуализации многомерных данных. Системы рассуждений на аналогиях. Деревья решений. Алгоритмы ограниченного перебора. Эволюционное программирование. Генетические алгоритмы. Нейронные сети.

Тема 4. Методы анализа потоков данных и ссылок и их применение в логистике

Потоковая модель данных и возможности ее применения в логистике. Выборка данных из потока. Фильтрация потоков. Подсчет различных элементов в потоке. Оценивание моментов. Подсчет единиц в окне. Затухающие окна. PageRank. Эффективное вычисление PageRank. Тематический PageRank. Ссылочный спам. Хабы и авторитетные страницы и их формализация.

Тема 5. Кластеризация и самоорганизующиеся карты Кохонена

Введение в методы кластеризации. Методы разбиения множества исследуемых объектов на однородные группы-кластеры. Иерархическая кластеризация: эффективность и альтернативные правила управления. Самоорганизующиеся карты Кохонена как разновидность нейросетевых алгоритмов. Возможности решения задач кластеризации и сегментирования в логистике с помощью специальных программных продуктов.

5.2. Учебно-тематический план

Таблица 2

№ п/ п	Наименование тем (разделов) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемо сти
		Всего	Контактная работа – Аудиторная работа			Самостоя тельная работа	
			Общая , в т.ч.:	Лекции	Семинары, практически е занятия		
1	Тема 1. Методы	21	6	2	4	15	Опрос в устной

	применения Data Mining и Knowledge discovery in databases в логистике.						и/или письменной форме, практико-ориентированные и ситуационные задания
2	Тема 2. Методы интеллектуального анализа больших данных в логистике.	21	6	2	4	15	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания
3	Тема 3. Классы программных средств интеллектуального анализа больших данных в логистике.	26	6	2	4	20	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания
4	Тема 4. Методы анализа потоков данных и ссылок и их применение в логистике	26	6	2	4	20	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания
5	Тема 5. Кластеризация и самоорганизующиеся карты Кохонена.	14	6	2	4	8	Опрос в устной и/или письменной форме,

							практико-ориентированные и ситуационные задания
	В целом по дисциплине	108	30	10	20	78	Согласно учебному плану: Домашнее творческое задание
	Итого в %	100	28	33	67	72	

5.3. Содержание семинаров, практических занятий

Таблица 3

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9	Формы проведения занятий
Тема 1. Методы применения Data Mining и Knowledge discovery in databases в логистике.	1. Data Mining в теории анализа больших данных. 2. Базы данных как элемент Data Mining. 3. Прикладная статистика как основа Data Mining. 4. Роль машинного обучения в формировании Data Mining. Рекомендуемые источники: раздел 8, №№ 1-5; раздел 9, №№ 1-34.	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания.
Тема 2. Методы интеллектуального анализа больших данных в логистике.	1. Интеллектуальный анализ и его функции в логистике. 2. Интеллектуальный анализ и его задачи в логистике. 3. Ассоциации как технология поиска зависимостей. 4. Последовательность как метод поиска зависимостей. 5. Классификация как метод выделения классов в логистике. 6. Кластеризация как метод выделения классов в логистике. Рекомендуемые источники: раздел 8, №№ 1-5; раздел 9, №№ 1-34.	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания.
Тема 3. Классы программных средств	1. Системы интеллектуального анализа данных и их категории в логистике. 2. Аналитические системы, имеющие	Опрос в устной и/или письменной форме,

интеллектуального анализа больших данных в логистике.	<p>предметно-ориентированную направленность.</p> <p>3. Применение статистических пакетов в логистике.</p> <p>4. Особенности применения программы StatSoft в логистике.</p> <p>5. Особенности применения программы SPSS в логистике.</p> <p>6. Программы визуализации многомерных данных и их применение в логистике.</p> <p>Рекомендуемые источники: раздел 8, №№ 1-5; раздел 9, №№ 1-34.</p>	практико-ориентированные и ситуационные задания.
Тема 4. Методы анализа потоков данных и ссылок и их применение в логистике.	<p>1. Возможности применения потоковой модели данных в логистике.</p> <p>2. Автоматизация процесса выборки данных их потока и особенности ее применения в логистике.</p> <p>3. Метод фильтрации потоков и его применение в логистике.</p> <p>4. Технология подсчета различных элементов в логистическом потоке.</p> <p>5. Технология оценивания моментов в логистике.</p> <p>6. Применение подсчета единиц в окне как способ анализа потоков данных.</p> <p>Рекомендуемые источники: раздел 8, №№ 1-5; раздел 9, №№ 1-34.</p>	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания.
Тема 5. Кластеризация и самоорганизующиеся карты Кохонена.	<p>1. Методы кластеризации и их применение в логистике.</p> <p>2. Кластеризация как метод разбиения множества исследуемых объектов на однородные группы-кластеры.</p> <p>3. Эффективность иерархической кластеризации в логистике.</p> <p>4. Альтернативные правила управления иерархической кластеризацией в логистике.</p> <p>Рекомендуемые источники: раздел 8, №№ 1-5; раздел 9, №№ 1-34.</p>	Опрос в устной и/или письменной форме, практико-ориентированные и ситуационные задания.

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Таблица 4

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Тема 1. Методы применения Data Mining и Knowledge discovery in databases в логистике.	<ol style="list-style-type: none"> 1. Методы стратегической работы с данными. 2. Применение статистического анализа в технологии Data Mining. 3. Применение деревьев решений в технологии Data Mining. 4. Нейросети, методы поиска логических правил, методы визуализации найденных закономерностей как технологии Data Mining. 	<ul style="list-style-type: none"> - работа с конспектом лекции; - работа с электронной библиотечной системой; - работа с информационно-образовательным порталом Финуниверситета; - подготовка к решению задач.
Тема 2. Методы интеллектуального анализа больших данных в логистике.	<ol style="list-style-type: none"> 1. Обобщение как метод описания концепций в логистике. 2. Распознавание как метод описания концепций в логистике. 3. Сравнение как метод описания концепций в логистике. 4. Обнаружение аномалий как отклонений от нормального логистического процесса. 5. Обнаружение выбросов и использование этой технологии обнаружения отклонений в логистике. 6. Анализ тенденций и его роль в развитии логистики. 7. Методы визуализации данных в логистике. 	<ul style="list-style-type: none"> - работа с конспектом лекции; - работа с электронной библиотечной системой; - работа с информационно-образовательным порталом Финуниверситета; - подготовка к решению задач.
Тема 3. Классы программных средств интеллектуального анализа больших данных в логистике.	<ol style="list-style-type: none"> 1. Возможности применения систем рассуждений на аналогиях в логистике. 2. Деревья решений как метод автоматического анализа больших массивов данных в логистике. 3. Особенности применения алгоритмов ограниченного перебора в логистике. 4. Метод эволюционного программирования в логистике. 5. Технологии включения генетических алгоритмов в логистические процессы 6. Нейронные сети в логистике. 	<ul style="list-style-type: none"> - работа с конспектом лекции; - работа с электронной библиотечной системой; - работа с информационно-образовательным порталом Финуниверситета;

		- подготовка к решению задач.
Тема 4. Методы анализа потоков данных и ссылок и их применение в логистике.	<ol style="list-style-type: none"> 1. Ссылки и их применение в логистике. 2. Затухающие окна как метод анализа потоков данных. 3. Применение PageRank в логистике. 4. Эффективное вычисление PageRank для логистических потоков. 5. Тематический PageRank в логистических процессах. 6. Ссылочный спам и методы борьбы с ним в логистике. 7. Формализация хабов в логистике. 8. Формализация авторитетных страниц в логистике. 	<ul style="list-style-type: none"> - работа с конспектом лекции; - работа с электронной библиотечной системой; - работа с информационно-образовательным порталом Финуниверситета; - подготовка к решению задач.
Тема 5. Кластеризация и самоорганизующиеся карты Кохонена.	<ol style="list-style-type: none"> 1. Особенности применения самоорганизующихся карт Кохонена в логистике. 2. Разновидности нейросетевых алгоритмов и их применение в логистике. 3. Методы решения задач кластеризации в логистике с помощью специальных программных продуктов. 4. Методы решения задач сегментирования в логистике с помощью специальных программных продуктов. 	<ul style="list-style-type: none"> - работа с конспектом лекции; - работа с электронной библиотечной системой; - работа с информационно-образовательным порталом Финуниверситета; - подготовка к решению задач.

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Текущий контроль успеваемости осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения ими самостоятельных работ. Основной *формой* текущего контроля знаний является домашнее творческое задание (ДТЗ).

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Департамента логистики и маркетинга.

Перечень примерных тем для выполнения домашнего творческого задания

1. Возможности применения Data Mining в логистике.
2. Теория баз данных как элемент Data Mining в логистике.
3. Теория прикладной статистики как элемент Data Mining в логистике.
4. Теория машинного обучения как элемент Data Mining в логистике.
5. Стратегии извлечения данных и их применение в логистике.
6. Статистический анализ как технология Data Mining.
7. Деревья решений как технология Data Mining.
8. Методы поиска логических правил как технология Data Mining.
9. Нейросети как технология Data Mining.
10. Методы визуализации найденных закономерностей как технология Data Mining.
11. Теоретические основы knowledge discovery in databases и возможности их применения в логистике.
12. Возможности применения интеллектуального анализа данных в логистике.
13. Поиск зависимостей (ассоциации, последовательности) как метод интеллектуального анализа данных в логистике.
14. Выделение классов (классификация, кластеризация) как метод интеллектуального анализа данных в логистике.
15. Описание концепций (обобщение, распознавание, сравнение) как метод интеллектуального анализа данных в логистике.
16. Обнаружение отклонений (обнаружение аномалий, выбросов; анализ тенденций) как метод интеллектуального анализа данных в логистике.
17. Визуализация данных как метод интеллектуального анализа данных в логистике.

18. Системы интеллектуального анализа данных и их применение в логистике.
19. Предметно-ориентированные аналитические системы и их применение в логистике.
20. Статистические пакеты и их применение в логистике.
21. Системы визуализации многомерных данных и их применение в логистике.
22. Системы рассуждений на аналогиях и их применение в логистике.
23. Деревья решений и их применение в логистике.
24. Алгоритмы ограниченного перебора и их применение в логистике.
25. Эволюционное программирование и его применение в логистике.
26. Генетические алгоритмы и их применение в логистике.
27. Нейронные сети и их применение в логистике.
28. Поточковая модель данных и возможности ее применения в логистике.
29. Анализ потоковых данных и его применение в логистике.
30. Анализ ссылок и его применение в логистике.
31. Методы и возможности применения кластеризации в логистике.
32. Методы и возможности применения самоорганизующихся карт Кохонена в логистике.

Пример задачи распознавания благонадежных и неблагонадежных заемщиков.

Исходные данные. В файле *credit.xls* собрано $C = 999$ записей о возвратах или невозвратах ранее выданных кредитов. Фрагмент файла приведен в следующей таблице:

Возраст	Пол	Состоит в браке	Иждивенцы	Доход	Опыт работы	Срок проживания	Недвижимость	Месячный платеж	Благонадежный заемщик
28	женский	Да	0	9000	9	7	0	3946	Нет
39	мужской	Да	1	13500	17	6	0	2460	Да
31	мужской	Нет	2	7000	11	3	0	3126	Нет
34	мужской	Нет	1	10200	15	2	41	3280	Да
46	женский	Да	2	8500	20	8	0	3348	Да
30	женский	Да	2	9500	12	30	35	4612	Нет
47	мужской	Нет	2	7900	14,5	6	67	2870	Нет
33	мужской	Нет	2	12600	15	23	0	2050	Да
22	мужской	Нет	0	34000	4	19	33	2562	Да
30	мужской	Да	1	33000	10	8	0	1948	Да
26	мужской	Да	1	13500	8	2	31	5535	Да
48	мужской	Нет	2	24000	22	40	40	1845	Да
37	мужской	Да	1	12500	15,5	6	18	2409	Да
27	мужской	Да	1	4000	9	6	0	4305	Нет
55	мужской	Да	1	5500	11,5	35	28	3382	Нет
39	женский	Нет	1	8000	6	8	0	4818	Нет
29	мужской	Да	1	11000	11	29	38	5228	Да
36	мужской	Нет	1	11600	18	13	0	4373	Да
32	мужской	Да	1	8500	14	28	0	2101	Да
35	женский	Нет	0	4500	16	33	0	1913	Нет
46	мужской	Да	1	9000	16	28	0	718	Да
22	мужской	Да	1	4000	3	20	21	6252	Нет
37	мужской	Да	2	29000	12	6	0	4612	Да
56	женский	Да	0	29000	19	36	34	4237	Да
34	мужской	Нет	1	26000	16	5	40	6662	Да
40	мужской	Нет	2	10500	20,5	2	25	3895	Нет
33	мужской	Нет	1	10700	15	3	24	4612	Нет
37	мужской	Да	2	35000	17,5	31	0	3485	Да
24	мужской	Да	0	11500	6	8	0	2972	Да
30	мужской	Да	2	15500	6	6	20	1435	Да
32	мужской	Нет	1	11300	14	28	0	6560	Нет
25	мужской	Нет	2	12500	7	17	0	1948	Да

В качестве 9 признаков описания заемщиков используются: возраст, пол, семейный статус, количество иждивенцев, подтвержденный документами месячный доход (в рублях), опыт работы (в годах), срок проживания в данном городе (в годах), рыночная стоимость недвижимости (в усл.ед.), ежемесячный платеж по кредиту (в рублях).

Обратите внимание, что 7 признаков из 9 — количественные (непрерывные). Оставшиеся 2 признака — Пол и «Состоит в браке» — дискретные (в них занесены нечисловые значения). Пропущенных данных в таблице нет.

Метка класса размещена в столбце J, где указано, был ли кредит возвращен вовремя (Нет или Да). Таким образом заданы 2 класса объектов: класс *A* (назовем его Дефолт) и класс *B* (назовем его Норма). Количество записей с определенным значением метки можно рассчитать в *Excel*, например, с помощью функции СЧЁТЕСЛИ. Всего в таблице приведены данные о $P = 507$ неблагонадежных заемщиках и о $N = 492$ благонадежных. В качестве упражнения рассчитайте, сколько клиентов состояло в браке на момент заполнения заявки на кредит.

Необходимо сопоставить компьютерный и истинный диагнозы и построить сопряженности, отражающую возможные соотношения между истинным и компьютерным диагнозом.

Решение:

Допустим, что с помощью некоторого метода были построены решающие правила, позволяющие диагностировать объекты, т.е. для каждой записи можно получить компьютерный диагноз Норма или Дефолт. Построим таблицу сопряженности, отражающую возможные соотношения между истинным и компьютерным диагнозом.

Таблица сопряженности

Истинный диагноз	Результат тестирования	
	Диагноз <i>D1</i> Дефолт	Диагноз <i>D2</i> Норма
Класс <i>A</i> Дефолт	<i>TP</i>	<i>FN</i>
Класс <i>B</i> Норма	<i>FP</i>	<i>TN</i>

В таблице использованы следующие сокращения:

- *TP* (сокращение от *true positive*) — истинноположительные диагнозы, т.е. число правильно поставленных диагнозов Дефолт;
- *TN* (*true negative*) — истинноотрицательные диагнозы, т.е. число правильно поставленных диагнозов Норма;
- *FP* (*false positive*) — ложноположительные диагнозы, т.е. число благонадежных клиентов, которым ошибочно рекомендовано не выдавать кредит;
- *FN* (*false negative*) — ложноотрицательные диагнозы, т.е. число случаев, когда кредит рекомендовано выдать, а клиент оказался неблагонадежным.

Так как наша первоочередная задача — выявить неблагонадежных клиентов, то «истинноположительным» в этом случае логично назвать правильный диагноз Дефолт.

Итак, разногласия между истинным и компьютерным диагнозами могут быть 2 типов:

- ложноположительный диагноз, называемый также *гипердиагностикой* или *ложной тревогой*;
- ложноотрицательный диагноз, называемый *гиподиагностикой* или *пропуском цели*.

Пример практико-ориентированного задания.

Для исходных данных, представленных в задаче распознавания благонадежных и неблагонадежных заемщиков, необходимо рассчитать

количественные характеристики, применимые для оценки качества отдельных решающих правил или их совокупности. При этом для оценки одного, отдельно рассматриваемого правила можно использовать показатели точности и полноты. Допустим, найдено правило R , позволяющее определять заемщиков, которые не вернут кредит (возможно, не всех, а только определенный процент из P имеющихся в базе данных). Пусть R позволило идентифицировать P_R потенциальных дефолтов и ошибочно отнесло к рискованным N_R заявок.

Задание:

1. Придумайте правило.
2. Рассчитайте полноту придуманного правила.
3. Рассчитайте точность придуманного правила.
4. Проведите аналогичные расчеты полноты и точности для следующего «правила»: «не выдавать кредит, если заемщик не состоит в браке и имеет иждивенцев». Не забывайте обновлять сводную таблицу для пересчета итоговых значений.
5. Придумайте «правило», отражающее ваши собственные представления о характерных особенностях благонадежных или неблагонадежных заемщиков. Оцените качество этого правила.

Пример ситуационного задания.

Для исходных данных, представленных в задаче распознавания благонадежных и неблагонадежных заемщиков, необходимо рассчитать чувствительность и специфичность.

В нашем примере *чувствительность* $Sens$ (*sensitivity*) — это доля невозвратов, правильно распознанных с помощью компьютерного правила, по отношению к общему числу невозвратов: $Sens = TP / (TP + FN)$.

Специфичность *Spec (specificity)* — доля благонадежных заемщиков, признанных таковыми и по результатам теста: $Spec = TN / (TN + FP)$.

Компьютерная система классификации может претендовать на внедрение в практику, если она обладает достаточно высокими значениями как чувствительности, так и специфичности.

Также для оценки эффективности таких интеллектуальных систем можно применять и показатель точности, т.е. отношение правильно поставленных диагнозов к общему числу объектов: $Acc = (TN + TP) / (P + N)$.

Проиллюстрируем эти определения. Реализуйте следующее решающее «правило»:

ЕСЛИ доход клиента >12000

И он является собственником объекта недвижимости

ТО кредит выдается

ИНАЧЕ

ЕСЛИ клиент старше 25 лет, состоит в браке и выплаты по кредиту не превысят 40% от месячного дохода

ТО кредит выдается

ИНАЧЕ отказать в кредите.

Рассчитайте показатели чувствительности и специфичности.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине содержится в разделе «2. Перечень планируемых результатов освоения образовательной

программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Таблица 5

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
<p><u>ПКН-2</u></p> <p>Способность применять современные методы и техники сбора, обработки и анализа данных, а также определения и прогнозирования основных социально-экономических показателей объектов управления</p>	<p>1. Разрабатывает методы, техники и инструментарий для анализа и прогнозирования тенденций и социально-экономических показателей</p>	<ul style="list-style-type: none"> • Знать: особенности анализа и прогнозирования тенденций и социально-экономических показателей и технико-методический инструментарий их разработки • Уметь: применять методы, техники и инструментарий для анализа и прогнозирования тенденций и социально-экономических показателей 	<p>Задание</p> <p>Примером предметно-ориентированных аналитических систем служат системы «технического анализа», получившие широкое распространение в области исследования финансовых рынков. Технический анализ представляет собой совокупность десятков методик прогнозирования динамики цен, формирования оптимальной структуры инвестиционного портфеля и т.п. Сделайте описание 3 программных продуктов, которые являются предметно-ориентированными аналитическими системами и могут применяться для анализа больших данных в логистике.</p> <p>Задание</p> <p>Среди наиболее мощных и распространенных статистических пакетов можно прежде всего назвать <i>SPSS</i>, <i>Statsoft</i> <i>STATISTICA</i> и <i>MathWorks MATLAB</i>. К недостаткам систем этого типа следует</p>

			<p>отнести:</p> <ul style="list-style-type: none"> • высокие требования к специальной подготовке пользователя, • перегруженность всевозможными функциями («тяжеловесность») • и немалую стоимость. <p>Укажите, в чем заключаются достоинства данных программных продуктов и опишите возможности их применения для анализа больших данных в логистике.</p>
	<p>2. Использует инструменты диагностики изменения состояния объектов управления на ранних стадиях в целях прогнозирования результатов их деятельности и предотвращения негативных последствий</p>	<ul style="list-style-type: none"> • Знать: возможные варианты состояния объектов управления на ранних стадиях, методы прогнозирования результатов их деятельности и способы предотвращения негативных последствий • Уметь: применять инструменты диагностики изменения состояния объектов управления на ранних стадиях в целях прогнозирования результатов их деятельности и предотвращения негативных последствий 	<p>Задание</p> <p>Идея систем рассуждений на аналогиях (<i>case based reasoning</i> — <i>CBR</i>) формулируется так: чтобы найти адекватное решение или сделать прогноз на будущее, следует найти в прошлом близкие аналоги текущей ситуации (прецеденты) и выбрать ответ, который был подтвержден как правильный.</p> <p>Что в данном случае содержит база знаний в <i>CBR</i>-системах: описание обобщенных ситуаций или сами конкретные ситуации? Ответ обоснуйте.</p> <p>Задание</p> <p>Системы, базирующиеся на принципах <i>CBR</i>, широко применяются, а среди присущих им недостатков</p>

			<p>указывается:</p> <ul style="list-style-type: none"> • возможный произвол при выборе меры «близости» по совокупности разнородных признаков; • не создается никаких моделей или правил, обобщающих предыдущий опыт, поэтому затруднительно сказать, на основе каких конкретно факторов (а не прецедентов) CBR системы строят свои ответы. <p>Укажите, в чем заключаются достоинства CBR-систем и опишите их возможности их применения для анализа больших данных в логистике.</p>
	<p>3. Владеет способностью анализировать проблемы финансово-экономического состояния организаций и прогнозировать их последствия.</p>	<ul style="list-style-type: none"> • Знать: методы проблемно-ориентированного анализа финансово-экономического состояния логистических организаций и механизмы оценки последствий диагностированного состояния • Уметь: применять анализ финансово-экономического состояния логистических организаций в целях повышения эффективности их функционирования 	<p>Задание</p> <p>Деревья решений — один из популярных подходов к задачам интеллектуального анализа данных. Они представляют собой иерархическую древовидную структуру, состоящую из классифицирующих правил типа «ЕСЛИ Предпосылка, ТО Следствие» (так называемые <i>if-then rules</i>). Укажите вид вопросов, на которые требуется последовательно ответить для принятия решения, к какому классу отнести конкретный объект или ситуацию? Приведите примеры таких вопросов.</p> <p>Задание</p>

			<p>Система предназначена для обнаружения информативных шаблонов в данных. Результат работы представляется в виде деревьев решений или списков правил «ЕСЛИ П, ТО С» (<i>if-then rules</i>). Система отличается весьма скромным интерфейсом и в этом смысле ее освоение не представляет существенных трудностей. Опишите возможности и недостатки данной системы для ее применения в целях анализа больших данных в логистике.</p>
	<p>4.Применяет интеллектуальные информационные технологии для повышения эффективности управления знаниями.</p>	<ul style="list-style-type: none"> • Знать: виды и типологию интеллектуальных информационных технологий, применяемых для повышения эффективности управления знаниями в логистике • Уметь: проводить оценку эффективности управления знаниями в логистике с применением интеллектуальных информационных технологий 	<p>Задание Алгоритмы ограниченного перебора были предложены в середине 1960-х годов М.М. Бонгардом для поиска логических закономерностей в данных. С тех пор они продемонстрировали свою эффективность при решении многих задач. Примером коммерчески распространяемого программного продукта, в котором реализованы алгоритмы ограниченного перебора, является система <i>WizWhy</i> компании <i>WizSoft</i>. Опишите возможности и недостатки данной системы для ее применения в целях анализа больших данных в логистике.</p> <p>Задание</p>

			<p>Генетические алгоритмы являются мощным средством решения разнообразных комбинаторных задач и задач оптимизации. Их идеология базируется на классическом дарвиновском понимании закономерностей существования биологических популяций, которые развиваются в течение нескольких поколений, подчиняясь законам естественного отбора и принципу «выживает наиболее приспособленный». Впервые подобный алгоритм был предложен в 1960-х гг. Дж. Холландом. Приведите примеры программных продуктов, в которых заложены основы генетического алгоритмирования. Оцените возможности их использования для проведения анализа больших данных в логистике.</p>
<p><u>ПК-3</u> способность разрабатывать экономические, финансовые и математические модели бизнес-процессов в логистике</p>	<p>1. Владеет различными методиками моделирования бизнес-процессов логистических компаний</p>	<ul style="list-style-type: none"> • Знать: способы и этапы моделирования бизнес-процессов, инициированных логистическими компаниями • Уметь: применять основные методы моделирования бизнес-процессов в логистических компаниях 	<p>Задание Data Mining — сравнительно молодая область исследований, сформировавшаяся в последние десять лет на пересечении теории баз данных, прикладной статистики и машинного обучения. Теоретические основы Data Mining еще только создаются, в то время как теория баз данных, статистика, машинное обучение развиваются уже давно.</p>

			<p>Определите время, в которое создавались основы теории баз данных, статистики, машинного обучения.</p> <p>Задание Термин «шаблон» можно определить, как правило R вида «ЕСЛИ объект обладает некоторым комплексом характеристик $\{A, B, C \dots\}$, ТО он относится к группе/классу G». При этом объекты принадлежат некоторой совокупности F, а правило R обладает достаточно высокой достоверностью (точностью). Какими еще критериями должно обладать правило R в информационном смысле?</p>
	<p>2. Применяет методы разработки экономических, финансовых и математических моделей для создания и развития логистического бизнеса.</p>	<ul style="list-style-type: none"> • Знать: возможности применения экономических, финансовых и математических моделей в логистическом бизнесе. • Уметь: применять методы разработки экономических, финансовых и математических моделей для создания и развития логистического бизнеса 	<p>Задание Задача выделения классов применительно к анализу клиентской базы логистических компаний означает, что имеющиеся клиенты должны быть сгруппированы в классы (группы), причем сами классы заранее известны. Известность самих классов (например, «постоянные клиенты») вовсе не означает, что уже известно и правило отнесения конкретных объектов к этим классам. Формальный подход основывается на выделении групп по имеющимся количественным показателям. Для использования представлений о предметной области используется так</p>

		<p>называемая смысловая группировка. Что означает данный подход и как его можно использовать в логистике?</p> <p>Задание Описание концепции — это тоже задача группировки логистических клиентов по классам. Но в этом случае определения классов заранее неизвестны. Это примыкает к обычной задаче кластеризации, когда группы должны быть выделены по критерию сходства объектов внутри групп и существенных различий между группами.</p> <p>Для каких логистических целей может использоваться данный подход?</p>
3. Разрабатывает предложения по повышению эффективности бизнес-процессов в логистике с применением цифровых технологий	<ul style="list-style-type: none"> ● Знать: цифровые технологии, применение которых направлено на повышение эффективности логистических бизнес-процессов ● Уметь: управлять разработкой эффективных экономических решений по рациональной организации бизнес-процессов в логистике на основе цифровых технологий 	<p>Задание Процедуры обнаружения отклонений позволяют выделить аномальные случаи или заметить новые тенденции. Аномалии в такой постановке задачи — это объекты, существенно отличающиеся от типичного шаблона. Аномалии (выбросы) могут быть обнаружены из анализа средних, стандартных отклонений, коэффициентов вариации (волатильности) и др. Как рассчитываются данные показатели?</p>

			<p>Задание Процесс обнаружения знаний в базах данных — итерационный процесс. Выделяются, в частности, такие его этапы:</p> <ul style="list-style-type: none"> • постановка задачи (формулировка целей); • выборка данных из общей базы или использование всей БД; • визуализация данных; • выбор признаков (характеристик) для дальнейшего анализа; • построение новых признаков и преобразование исходных данных; • применение вычислительных методов интеллектуального анализа; • интерпретация полученных результатов, структурирование знаний; • при необходимости уточнение постановки задачи и повтор этапов. <p>Приведите практические примеры этапов обнаружения знаний в базе данных логистической компании.</p>
--	--	--	---

Примерный перечень вопросов к зачёту (устная форма проведения зачета):

1. Теоретические основы Data Mining.

2. Теория баз данных, прикладной статистики и машинного обучения.
3. Стратегии извлечения данных.
4. Технологии Data Mining: статистический анализ, деревья решений, нейросети, методы поиска логических правил, методы визуализации найденных закономерностей и др.
5. Теоретические основы knowledge discovery in databases «обнаружение знаний в базах данных».
6. Задачи интеллектуального анализа данных в логистике.
7. Поиск зависимостей (ассоциации, последовательности).
8. Выделение классов (классификация, кластеризация) в логистике.
9. Описание концепций (обобщение, распознавание, сравнение).
10. Обнаружение отклонений (обнаружение аномалий, выбросов; анализ тенденций).
11. Визуализация данных.
12. Категории систем интеллектуального анализа данных в логистике.
13. Предметно-ориентированные аналитические системы.
14. Статистические пакеты.
15. Системы визуализации многомерных данных.
16. Системы рассуждений на аналогиях.
17. Деревья решений.
18. Алгоритмы ограниченного перебора.
19. Эволюционное программирование.
20. Генетические алгоритмы.
21. Нейронные сети.
22. Поточная модель данных и возможности ее применения в логистике.
23. Выборка данных из потока.
24. Фильтрация потоков.

25. Подсчет различных элементов в потоке.
26. Оценивание моментов.
27. Подсчет единиц в окне.
28. Затухающие окна.
29. PageRank.
30. Эффективное вычисление PageRank.
31. Тематический PageRank.
32. Ссылочный спам.
33. Хабы и авторитетные страницы и их формализация.
34. Введение в методы кластеризации.
35. Методы разбиения множества исследуемых объектов на однородные группы-кластеры.
36. Иерархическая кластеризация: эффективность и альтернативные правила управления.
37. Самоорганизующиеся карты Кохонена как разновидность нейросетевых алгоритмов.
38. Возможности решения задач кластеризации и сегментирования в логистике с помощью специальных программных продуктов.

Методические материалы, определяющие процедуры оценивания знаний, умений и навыков

Приказ от 23.03.2017 №0557/о «Об утверждении Положения о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся по программам бакалавриата и магистратуры в Финансовом университете» и приказы филиалов по данному вопросу.

**8. Перечень основной и дополнительной учебной литературы,
необходимой для освоения дисциплины**

Нормативно-правовые акты:

1. Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 N 149-ФЗ

2. Федеральный закон «О защите персональных данных» от 27 июля 2006 г. № 152-ФЗ

Основная литература:

3. Логистика в цифровой экономике: тенденции и векторы развития : монография / Д. В. Швандар, А. А. Арский, Г. П. Быкова [и др.] ; под ред. И. А. Меркулиной, Ф. Д. Венде. — Москва : КноРус, 2023. — 210 с. — ЭБС BOOK.ru. — URL: <https://book.ru/book/946344> (дата обращения: 28.12.2022). — Текст : электронный.

Дополнительная литература:

4. Катаргин Н. В. Анализ и моделирование логистических систем: учебник / Н. В. Катаргин, О. Н. Ларин, Ф. Д. Венде; Финуниверситет - Санкт-Петербург [и др.]: Лань, 2021. - 246 с. - Текст : непосредственный. - То же. — ЭБС Лань. - URL: <https://e.lanbook.com/book/162375>; URL: <https://e.lanbook.com/book/179155> (дата обращения: 28.12.2022). - Текст : электронный.

5. Цифровая трансформация: IoT, AI, VR, Big Data / Digital Transformation: IoT, AI, VR, Big Data : сборник докладов XII международной студенческой научно-практической конференции / отв. за вып. М. А. Иванова. - Москва : Дело (РАНХиГС), 2019. - 256 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1405779> (дата обращения: 28.12.2022). - Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронные ресурсы БИК:

- Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
- Электронно-библиотечная система BOOK.RU <http://www.book.ru>
- Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
- Электронно-библиотечная система Znanium <http://www.znanium.com>
- Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
- Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
- Справочно-образовательная система Актион 360 <https://action360.ru/>
- Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
- Интернет-библиотека СМИ Public.Ru <https://public.ru/>
- Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
- Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
- Национальная электронная библиотека <http://нэб.рф/>
- Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
- Ресурсы информационно-аналитического агентства по финансовым рынкам Cbonds.ru <https://cbonds.ru/>
- СПАРК <https://spark-interfax.ru/>
- STATISTA <https://www.statista.com/>
- Academic Reference <http://ar.cnki.net/ACADREF>
- Пакет баз данных компании EBSCO Publishing, крупнейшего агрегатора научных ресурсов ведущих издательств мира <http://search.ebscohost.com>
- Henry Stewart Talks: Библиотека Онлайн Лекций по Бизнесу и Маркетингу <https://hstalks.com/business/>

- Электронная коллекция книг издательства Springer: Springer eBooks
<http://link.springer.com/>
- Электронные продукты издательства Elsevier
<http://www.sciencedirect.com>
- Emerald: Management eJournal Portfolio <https://www.emerald.com/insight/>
- JSTOR. Arts & Sciences I Collection <https://www.jstor.org/>
- Библиотека электронных публикаций Организации экономического сотрудничества и развития OECD iLibrary <https://www.oecd-ilibrary.org/>
- Scopus <https://www.scopus.com>
- Видеотека учебных фильмов «Решение» (тематические коллекции «Менеджмент», «Маркетинг. Коммерция. Логистика», «Юриспруденция», «Управление персоналом», «Психология управления» <http://eduvideo.online/>
- База данных научных журналов издательства Wiley
<https://onlinelibrary.wiley.com/>
- Цифровой архив научных журналов: <http://arch.neicon.ru/xmlui/>
 - Annual Reviews
 - Cambridge University Press
 - The Institute of Physics (IOP) Publishing
 - Nature
 - Oxford University Press
 - Royal Society of Chemistry
 - SAGE Publications
 - Science
 - Taylor & Francis Group

10. Методические указания для обучающихся по освоению дисциплины

Студентам при подготовке следует использовать нормативные документы Финансового университета, а именно:

Методические рекомендации по планированию и организации внеаудиторной самостоятельной работы студентов по образовательным программам бакалавриата и магистратуры в Финансовом университете, утвержденные приказом Финуниверситета от 11.05.2021 г. № 1040 (см. сайт Финансового Университета: на главной странице раздел «Наш университет»; далее «Единая правовая база Финуниверситета»), использовать методические рекомендации департамента.

Методические рекомендации по решению задач распознавания благонадежных и неблагонадежных заемщиков.

Целью решения задач является освоение базовых методик оценки эффективности компьютерных систем распознавания.

Компьютерные системы распознавания создаются для решения сложных практических задач, в которых необходимо правильно отнести тестируемый объект к определенному классу из нескольких, с которыми умеет работать данная система. Такие задачи называются задачами классификации, и в самом простом случае в них различаются 2 класса (обычно обозначаемых A и B , или A и $неA$). Например, медицинские диагностические системы позволяют получить ответ на вопрос, болен или здоров обследуемый пациент; банковские системы позволяют принять решение выдавать или не выдавать кредит; системы аварийной сигнализации предупреждают о наступлении опасной ситуации и т.д.

Алгоритмическую основу подобных компьютерных систем составляет так называемое **решающее правило**, на вход которого подаются значения признаков анализируемого объекта, а результатом применения правила является ответ, к классу A или к классу B следует данный объект отнести. Иногда такой ответ называется **компьютерным диагнозом** (по аналогии с диагнозом медицинским).

В некоторых системах допускается ответ, означающий отказ от классификации конкретного объекта («нельзя отнести ни к A , ни к B »). В других системах такой отказ неприемлем и непременно должен быть получен однозначный ответ (либо A , либо B).

Формально решающее правило может быть представлено в виде

ЕСЛИ $\Psi(X)$, ТО $X \rightarrow A$, ИНАЧЕ $X \rightarrow B$.

При этом такое обобщенное решающее правило Ψ может состоять из комплекса отдельных частных правил ϕ_i , каждое из которых оказывается полезным для классификации какой-то части объектов X .

Для построения решающего правила необходимо пройти следующие основные этапы:

- этап сбора данных (примеров) и формирования материала обучения;
- этап обучения с использованием какого-либо вычислительного метода;
- этап тестирования полученного правила на материале контроля (контрольных примерах).

Тестирование подразумевает, что имеется возможность сравнить на контрольных объектах результат компьютерного распознавания и истинный диагноз. Например, собрана достаточная статистика по невозврату выданных кредитов. Про каждого заемщика уже достоверно известно — выплатил он кредит или нет. Подавая на вход решающего правила формализованные данные заемщика (пол, возраст, размер дохода и его источники, кредитную историю и т.д.), можно сравнить рекомендацию компьютерной системы («выдавать» или «не выдавать») с известным результатом.

Весьма вероятно, что для части контрольных примеров компьютерный диагноз будет не совпадать с истинным: компьютерная рекомендация

«выдавать» для некредитоспособного и рекомендация «не выдавать» для надежного клиента.

Оценка эффективности решающего правила будет объективна, если контрольные примеры, на которых правило тестируется, не входили в материал обучения, использовавшийся при построении этого правила. Когда доступных для анализа объектов слишком мало, аналитик вынужден проводить тестирование на тех же объектах, что и обучение. Следует помнить, что при этом, вероятнее всего, будет получена завышенная оценка качества правила и при практическом применении доля ошибок компьютерной системы почти наверняка возрастет.

В настоящей работе этап сбора данных будет пропущен (необходимые данные уже подготовлены), а основное внимание будет уделено 2 показателям качества построенного решающего правила —чувствительности и специфичности. Для упрощения выкладок и вычислений будет использоваться вариант, когда материал контроля и материал обучения совпадают.

Методические рекомендации по решению практико-ориентированных заданий

Точность правила *Acc* (*accuracy*) — это доля случаев, когда правило дает верный результат, относительно всех случаев его применения. В нашем примере точность вычисляется как количество оправданно отклоненных заявок, деленное на общее количество случаев, когда правило **R** выдавало рекомендацию отклонить заявку: $Acc = P_R / (P_R + N_R)$.

Полнота правила *Comp* (*completeness*) — это доля случаев, когда правило подтверждается, относительно количества объектов в интересующем нас классе. Так как правило **R** прогнозирует невозвраты, то

его полнота равна отношению верных предсказаний дефолта к их общему числу в базе: $Comp = P_R / P$.

Скопируйте лист с исходными данными. На созданной копии определите следующее «правило»: «если доход клиента меньше среднего и при этом клиент не имеет в собственности недвижимости, то кредит выдавать не следует». Формализуйте это правило в виде формулы в столбце К, который озаглавьте Прогноз. Обратите внимание, что данное правило не включает части ИНАЧЕ и, следовательно, из факта, что доход клиента больше среднего или клиент имеет недвижимость, вообще говоря, не предполагается делать вывод о его кредитоспособности.

Для удобства оценки правила постройте сводную таблицу.

«Количество по полю Пол» позволяет получить число клиентов. Прогноз разместим в полях столбцов, а поле Благонадежный заемщик — в полях строк.

Количество по полю Пол	Прогноз		
Благонадежный заемщик	0	Нет	Общий итог
Да	414	78	492
Нет	229	278	507
Общий итог	643	356	999

Рис. 1. Сводная таблица для оценки эффективности правила

Рассчитаем полноту придуманного правила. Оно описывает 278 из 507 случаев невозвратов, т.е. полнота правила составляет $Comp = 54.8\%$.

Рассчитаем точность правила. Оно применялось 356 раз, из них дало верный результат в 278 случаях. Таким образом $Acc = 78.1\%$.

Методические рекомендации по решению ситуационных заданий

После необходимых вычислений постройте сводную таблицу (см. рис. 2), из которой можно получить следующие

Количество по полю Пол	Прогноз		
Благонадежный заемщик	Нет	Да	Общий итог
Нет	403	104	507
Да	182	310	492
Общий итог	585	414	999

Рис. 2. Оценки чувствительности и специфичности

значения показателей: $TP = 403$;
 $FN = 104$; $FP = 182$; $TN = 310$.

Рассчитайте характеристики решающего правила:

- чувствительность $Sens = TP / (TP + FN)$;
- специфичность $Spec = TN / (TN + FP)$;
- точность $Acc = (TN + TP) / (TN + TP + FN + FP)$.

Сохраните (скопируйте) полученную таблицу, она потребуется в дальнейшем.

Сравните этот результат с выставлением компьютерного диагноза с помощью генератора случайных чисел. Используйте функцию Excel СЛЧИС, которая возвращает равномерно распределенное случайное число в интервале $[0; 1)$. Положим, что если функция вернула значение большее или равное 0.5, то будем трактовать его как диагноз Нет, а если меньше — то как Да.

Возможный вариант расчета представлен на рис. 3. Для него:

$TP = 260$; $FN = 247$; $FP = 249$;
 $TN = 243$. Отсюда $Sens = 51.2\%$;
 $Spec = 49.4\%$; $Acc = 50.4\%$.

Количество по полю Пол	Прогноз		
Благонадежный заемщик	Нет	Да	Общий итог
Нет	260	247	507
Да	249	243	492
Общий итог	509	490	999

Рис. 3. Сводная таблица для

Итак, в случае распознавания двух классов даже случайное угадывание дает значения чувствительности, специфичности и точности около 50% (при примерно равном количестве объектов в обоих классах). Таким образом, специфичность в 63%, полученная на примере с рис. ПР12.2, ненамного выше, чем специфичность при случайном угадывании.

Допустим, что при реализации компьютерного распознавания допущена ошибка и теперь компьютерный тест любой объект относит в группу Дефолт. Результаты такого «тестирования» представлены на рис. 4.

Из значений $TP = 507$;
 $FN = 0$; $FP = 492$; $TN = 0$ можно
 получить, что $Sens = 100\%$;
 $Spec = 0\%$; $Acc = 50.8\%$.

Количество по полю Пол	Прогноз		
Благонадежный заемщик	Нет	0	Общий итог
Нет	507		507
Да	492		492
Общий итог	999		999

Рис. 4. Сводная таблица при ошибке в реализации компьютерного распознавания

Достигнуто стопроцентное значение чувствительности, т.е. выявляются все невозвраты, что отнюдь не является достаточным показателем качества этого «правила». Причина в том, что оно обладает нулевой специфичностью, т.е. кредиты при следовании такому правилу вообще никому выданы не будут.

В большинстве практических задач с увеличением чувствительности снижается специфичность и наоборот. Поэтому перед исследователем стоит нетривиальная задача нахождения «золотой середины», т.е. правил, одновременно характеризующихся достаточно высокими значениями как чувствительности, так и специфичности.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

- Windows Microsoft Office;
- Антивирус Kaspersky.

11.2. Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Консультант Плюс»;
2. Информационно-правовая система «Гарант».

11.3. Сертифицированные программные и аппаратные средства защиты информации

Не предусмотрено.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Для осуществления образовательного процесса по дисциплине «Методы анализа больших данных в логистике», необходимо:

1. Учебно-лабораторное оборудование:

- персональный компьютер.
- проектор.

2. Программные, технические и электронные средства обучения и контроля знаний студентов, размещенные на портале Финансового университета и доступные для использования в точках удаленного доступа и/или в помещениях Университета (электронная библиотека, программы для компьютерного тестирования, видеолекции, учебно-методические материалы и др.).