

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«КАЛИНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ»

**М. А. Романов, А. Б. Тристанов**

## **ТЕХНОЛОГИИ DATA MINING**

Учебно-методическое пособие по изучению дисциплины  
для студентов магистратуры по направлению подготовки  
09.04.01 – Информатика и вычислительная техника

Калининград  
Издательство ФГБОУ ВО «КГТУ»  
2022

Рецензент:

кандидат физико-математических наук, исполняющий обязанности  
заведующего кафедрой прикладной математики и информационных технологий  
ФГБОУ ВО «Калининградский государственный технический университет»  
Алексей Иванович Руденко

**Романов, М. А.**

Технологии Data Mining : учебно-методическое пособие по изучению  
дисциплины для студентов магистратуры по направлению подготовки 09.04.01 –  
Информатика и вычислительная техника / М. А. Романов, А. Б. Тристанов. –  
Калининград : Издательство ФГБОУ ВО «КГТУ». – 2022. – 18 с.

В учебно-методическом пособии приведен тематический план изучения  
дисциплины. Представлены методические указания по изучению дисциплины.  
Даны рекомендации по подготовке к промежуточной аттестации и критерии  
оценивания. Пособие подготовлено в соответствии с требованиями  
утвержденной рабочей программы модуля «Общепрофессиональный модуль  
(В)» по дисциплине «Технологии Data Mining» направления подготовки 09.04.01  
– Информатика и вычислительная техника.

Табл. 1, список лит. – 8 наименований

Учебно-методическое пособие рассмотрено и одобрено в качестве  
локального электронного методического материала на заседании кафедры  
прикладной математики и информационных технологий Института цифровых  
технологий ФГБОУ ВО «Калининградский государственный технический  
университет» 2 ноября 2022 г., протокол № 11.

Учебно-методическое пособие по изучению дисциплины рекомендовано к  
использованию в учебном процессе в качестве локального электронного  
методического материала методической комиссией Института цифровых  
технологий 3 ноября 2022 г., протокол № 8.

© Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Калининградский государственный технический  
университет», 2022 г.

© Романов М. А., Тристанов А. Б., 2022 г.

## ОГЛАВЛЕНИЕ

1. ВВЕДЕНИЕ .....	4
2. ТЕМАТИЧЕСКИЙ ПЛАН.....	5
3. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ И УКАЗАНИЯ К ИЗУЧЕНИЮ .....	6
3.1. Раздел 1. Введение в Data Mining .....	6
3.2. Раздел 2. Предобработка данных.....	7
3.3. Раздел 3. Обучение с учителем.....	8
3.4. Раздел 4. Обучение без учителя.....	10
3.5. Раздел 5. Искусственные нейронные сети.....	13
4. ТРЕБОВАНИЯ К АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ .....	15
4.1. Текущая аттестация.....	15
4.2. Условие получения положительной оценки .....	15
5. ЗАКЛЮЧЕНИЕ .....	16
6. ЛИТЕРАТУРА .....	16

## 1. ВВЕДЕНИЕ

Целью освоения дисциплины «Технологии Data Mining» является формирование у студентов целостного представления о технологии интеллектуального анализа больших наборов данных, а также знаний и умений в предварительной обработке данных; знакомство с базовыми концепциями технологии и структурой отдельных алгоритмов, а также с алгоритмами обнаружения новых, нетривиальных, практически полезных знаний в больших наборах данных.

В результате изучения дисциплины магистранты будут знать содержание технологии Data Mining и основные методы; понимать основные проблемы, возникающие при анализе больших данных, и знать пути их решения в рамках технологии Data Mining; выбирать методы интеллектуального анализа, исходя из практической задачи; владеть навыками анализа данных различной природы с использованием современных инструментальных средств.

Обучение проводится с использованием академической программы "Loginom Academic".

В предлагаемом пособии представлен тематический план, содержащий перечень изучаемых тем, обязательных лабораторных занятий, мероприятий текущей аттестации и отводимое на них аудиторное время (занятия в соответствии с расписанием) и на самостоятельную работу. При формировании личного образовательного плана на семестр обучающемуся следует оценивать рекомендуемое время на изучение дисциплины и возможность больших временных затрат на выполнение отдельных заданий или проработку отдельных тем.

В разделе «Содержание дисциплины» приведены подробные сведения о вопросах, рассматриваемых в данном курсе. Представлены методические рекомендации преподавателя для самостоятельной работы магистранта. Каждая тема включает ссылку на литературу (или иной информационный ресурс), а также контрольные вопросы для самопроверки и тесты для самодиагностики по изученной теме.

Раздел «Текущая аттестация» содержит описание обязательных мероприятий контроля самостоятельной работы и усвоения разделов или отдельных тем дисциплины. Изложены требования к промежуточной аттестации, проходящей в форме зачета.

Помимо данного пособия, магистрантам следует использовать материалы, размещенные в соответствующем разделе курса по дисциплине «Технологии Data Mining» в ЭИОС.

## 2. ТЕМАТИЧЕСКИЙ ПЛАН

	Раздел (модуль) дисциплины	Тема	Объем аудиторной работы, ч	Объем самостоятельной работы, ч
		<b>Лекции</b>		
1.1	Раздел 1. Введение в Data Mining	Тема 1. Введение в Data Mining	4	3
2.1	Раздел 2. Предобработка данных	Тема 2. Предобработка данных	4	7,5
3.1	Раздел 3. Обучение с учителем	Тема 3. Линейная регрессия	4	19,2
3.2		Тема 4. Классификация	4	10
4.1	Раздел 4. Обучение без учителя	Тема 5. Кластеризация	4	4
4.2		Тема 6. Ассоциативные правила	4	4
5	Раздел 5. Искусственные нейронные сети	Тема 7. Искусственные нейронные сети	6	13
			<b>30</b>	<b>60,7</b>
		<b>Лабораторные занятия</b>		
	Раздел 1. Введение в Data Mining	Лабораторная работа № 1. Введение в Data Mining	2	3
	Раздел 2. Предобработка данных	Лабораторная работа № 2. Предподготовка данных	8	4
	Раздел 3. Обучение с учителем	Лабораторная работа № 3. Регрессия	9	4
	Раздел 4. Обучение без учителя	Лабораторная работа № 4. Кластеризация	3	3
		Лабораторная работа № 5. Ассоциативные правила	3	3
	Раздел 5. Искусственные нейронные сети	Лабораторная работа № 6. Искусственные нейронные сети	5	4,15
			<b>30</b>	<b>21,15</b>
		<b>Рубежный (текущий) и итоговый контроль</b>		
		Итоговый контроль (зачет)	-	-
			<b>0</b>	<b>0</b>
		<b>Всего</b>	<b>60</b>	<b>81,85</b>

### 3. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ И УКАЗАНИЯ К ИЗУЧЕНИЮ

#### 3.1. Раздел 1. Введение в Data Mining

*Перечень изучаемых вопросов:*

Цели и задачи дисциплины. Место дисциплины в структуре образовательной программы. Планируемые результаты освоения дисциплины. Понятие анализа данных. Задачи анализа данных. Методы анализа данных. Принципы анализа данных. Инструменты анализа данных. Области применения анализа данных.

*Методические указания к изучению:*

Data Mining – это анализ больших объемов данных для выявления скрытых закономерностей, ассоциаций и шаблонов, которые можно использовать для принятия решений в различных областях, включая бизнес, медицину, науку и технологии. В настоящее время Data Mining является одной из ключевых областей в машинном обучении и искусственном интеллекте. Однако работа с большими объемами данных может быть сложной задачей. Для облегчения этого процесса созданы инструменты Data Mining. Цель этой технологии заключается не только в извлечении информации из больших объемов данных, но и в преобразовании этой информации в знания, которые можно использовать для принятия решений и определения стратегии. Data Mining используется для различных задач, таких как прогнозирование спроса на товары, выявление болезней и их причин, анализ социальных сетей и многих других. В данном разделе студенты познакомятся с основными понятиями и методами Data Mining, а также узнают, как их можно применять на практике. Они изучат различные методы Data Mining, включая кластерный анализ, классификацию и ассоциативные правила, и узнают, как эти методы могут быть использованы для решения практических задач. Результатом обучения студентов станет умение работать с большими объемами данных, извлекать из них знания и применять полученные знания для принятия решений в различных областях.

*Литература:*

В предлагаемой литературе [1, 2, 3, 4, 5, 8] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

*Контрольные вопросы:*

1. Что такое анализ данных?
2. Какие задачи решает анализ данных?
3. Какие методы используются в анализе данных?
4. Какие принципы лежат в основе анализа данных?

5. Какие инструменты используются в анализе данных?
6. Какие этапы включают в себя процесс анализа данных?
7. Какие проблемы могут возникнуть в процессе анализа данных?
8. Какие области применения анализа данных существуют?
9. Какие преимущества и недостатки имеет использование анализа данных?
10. Какие требования к данным для использования анализа данных?

### **3.2. Раздел 2. Предобработка данных**

*Перечень изучаемых вопросов:*

Очистка данных. Преобразование данных. Интеграция данных. Разделение данных. Определение типов данных. Обработка отсутствующих значений. Обработка выбросов. Масштабирование данных. Нормализация данных. Кодирование категориальных признаков. Слияние данных. Удаление дубликатов. Индексирование данных. Разделение на обучающую и тестовую выборки. Балансировка классов. Подготовка признаков для моделирования. Уменьшение размерности данных. Визуализация данных. Оценка качества данных. Сохранение предобработанных данных.

*Методические указания к изучению:*

Предобработка данных является одним из самых важных этапов в анализе данных и машинном обучении. Ее цель – подготовить данные для дальнейшей работы с ними, такой как анализ, обработка и использование в моделях машинного обучения. Основная задача предобработки данных заключается в улучшении качества данных и устранении ошибок, которые могут повлиять на результаты анализа. Данные, полученные из разных источников, могут содержать ошибки, пропуски, выбросы и другие аномалии, которые могут исказить результаты анализа. Предобработка данных включает в себя несколько этапов, таких как очистка данных, заполнение пропущенных значений, обработка выбросов и шумов, нормализация данных и масштабирование. Правильная предобработка данных может существенно повысить качество модели и ее точность. Качество и точность модели зависят от качества данных, на которых она обучается. Поэтому, если данные некачественны или содержат ошибки, модель также будет некачественной и работать неправильно. В данном разделе студенты рассмотрят основные методы предобработки данных, включая очистку данных, обработку выбросов и шумов, заполнение пропущенных значений, нормализацию и масштабирование данных. Также будут рассмотрены примеры использования этих методов на практике, чтобы показать, как они могут быть применены для улучшения качества данных и повышения точности модели. Применение правильной предобработки данных является одним из

ключевых факторов, влияющих на успех проектов по анализу данных и машинному обучению.

*Литература:*

В предлагаемой литературе [1, 2, 3, 4, 5, 8] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

*Контрольные вопросы:*

1. Что такое предобработка данных и почему она важна?
2. Какие этапы включает в себя предобработка данных?
3. Какие проблемы могут возникнуть в процессе предобработки данных?
4. Какие методы используются для очистки данных?
5. Какие методы используются для преобразования данных?
6. Какие методы используются для интеграции данных?
7. Какие методы используются для обработки отсутствующих значений?
8. Какие методы используются для обработки выбросов?
9. Какие методы используются для масштабирования данных?
10. Какие методы используются для визуализации данных?

*Рекомендуемая литература по разделу 2/теме 2.1:*

В предлагаемой литературе [1, 2, 3, 4, 5, 8] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

### **3.3. Раздел 3. Обучение с учителем**

#### **Тема 3.1. Линейная регрессия**

*Перечень изучаемых вопросов:*

Линейная регрессия. Модель линейной регрессии. Коэффициенты линейной регрессии. Метод наименьших квадратов. Регрессионный анализ. Критерий оптимальности. Предсказание. Множественная линейная регрессия. Регуляризация. Лассо-регрессия. Ридж-регрессия. Эластичная сеть. Кросс-валидация. Статистическая значимость. Коэффициент детерминации. Остатки. Анализ выбросов. Интерпретация результатов. Применение линейной регрессии в реальных задачах.

*Методические указания к изучению:*

Линейная регрессия – это один из наиболее популярных методов в статистике и машинном обучении, который позволяет моделировать зависимости между двумя переменными. Этот метод широко используется в различных областях, таких как экономика, финансы, бизнес-аналитика,

медицина и другие. С помощью линейной регрессии можно прогнозировать будущие результаты и описывать прошлые данные. Линейная регрессия – это простой, но мощный инструмент, который может помочь в решении различных задач в анализе данных. Он может быть использован для прогнозирования будущих результатов и для описания прошлых данных. Он также может быть использован для определения того, насколько изменение независимой переменной влияет на зависимую переменную. В данном разделе рассмотрены основы линейной регрессии, процесс построения модели и методы оценки ее качества. Студентам также будут показаны примеры ее применения на практике в различных областях, чтобы показать, как регрессия может быть применена для решения реальных задач. В результате они смогут лучше понимать, как использовать этот метод для анализа данных и как он может помочь им в работе.

#### *Литература:*

В предлагаемой литературе [1, 3, 5, 7, 8] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

#### *Контрольные вопросы:*

1. Что такое линейная регрессия и как она используется в анализе данных?
2. Каким образом происходит построение линейной регрессии?
3. Как можно оценить качество построенной модели линейной регрессии?
4. Какие основные предположения лежат в основе линейной регрессии?
5. Какие методы используются для выбора оптимальных параметров модели линейной регрессии?
6. Какие дополнительные регуляризационные методы могут применяться к линейной регрессии?
7. Какие методы используются для проверки значимости признаков в модели линейной регрессии?
8. Какие методы можно использовать для улучшения качества модели линейной регрессии?
9. Каким образом можно интерпретировать результаты, полученные с помощью линейной регрессии?
10. Какие области применения линейной регрессии существуют?

### Тема 3.2. Классификация

#### *Перечень изучаемых вопросов:*

Классификация. Класс. Объект. Признак. Алгоритм. Кластеризация. Разбиение. Дерево решений. Нейронные сети. Метод опорных векторов. Метрики качества. Переобучение. Регуляризация. Гиперпараметры. Отбор признаков. Кросс-валидация. Модель. Точность классификации.

*Методические указания к изучению:*

Классификация – один из наиболее важных методов анализа данных, широко используемый в различных областях, таких как медицина, финансы, бизнес-аналитика и другие. С помощью классификации можно разделить данные на группы схожих характеристик и свойств, что позволяет делать прогнозы и выводы на основе имеющихся данных. В этом разделе студенты смогут рассмотреть основы классификации, включая процесс построения модели и методы оценки ее качества. Для понимания будут предоставлены примеры ее применения на практике и объяснено, как использование правильной предобработки данных может повысить точность модели и ее качество. Кроме того, студенты рассмотрят различные методы классификации, которые могут быть использованы для более сложных наборов данных.

*Литература:*

В предлагаемой литературе [1, 2, 3, 7] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

*Контрольные вопросы:*

1. Что такое классификация и какие задачи она решает?
2. Какие основные типы классификации существуют?
3. Что такое объекты и признаки в контексте классификации?
4. Какие алгоритмы используются для классификации и как они работают?
5. Что такое кластеризация и как она отличается от классификации?
6. Что такое дерево решений и как оно используется в классификации?
7. Что такое нейронные сети и как они используются в классификации?
8. Что такое метод опорных векторов (SVM) и как он используется в классификации?
9. Что такое переобучение и как его избежать?
10. Как оценивается качество классификации и какие метрики используются для этого?

### **3.4. Раздел 4. Обучение без учителя**

#### **Тема 4.1. Кластеризация**

### *Перечень изучаемых вопросов:*

Кластеризация. Кластер. Разбиение. Иерархическая кластеризация. Алгоритм k-средних. EM-алгоритм. Алгоритм DBSCAN. Агломеративная кластеризация. Спектральная кластеризация. Кластеризация больших данных. Метрики качества кластеризации. Определение оптимального количества кластеров. Проблемы кластеризации и их решения. Применение кластеризации в машинном обучении. Применение кластеризации на практике в различных областях.

### *Методические указания к изучению:*

Кластеризация – это один из ключевых методов машинного обучения, который активно исследуется в науке о данных. Кластеризация – это процесс разделения набора данных на группы (кластеры) таким образом, чтобы объекты внутри кластера были максимально похожи между собой, а объекты между кластерами были максимально различными. Одним из основных преимуществ кластеризации является возможность обнаружения скрытых структур в данных, которые могут быть незаметными при первом взгляде. Также кластеризация может использоваться для обнаружения аномалий в данных, что полезно в таких областях, как банковское дело, медицина и транспорт. Хотя кластеризация является широко используемым методом, она также имеет ограничения. Например, выбор правильного числа кластеров может быть сложной задачей, и неправильный выбор может привести к неправильной интерпретации данных. Кроме того, некоторые алгоритмы кластеризации могут быть чувствительны к начальным значениям параметров и могут давать различные результаты при каждом запуске. В этом разделе будут рассмотрены основные алгоритмы кластеризации, такие как K-средних, DBSCAN и иерархическая кластеризация, а также их применения в различных областях, включая обработку изображений, биологию и маркетинг.

### *Литература:*

В предлагаемой литературе [1, 2, 4, 5, 6, 8] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

### *Контрольные вопросы:*

1. Что такое кластеризация и какую задачу она решает?
2. Какие методы кластеризации существуют и как они работают?
3. Как выбрать оптимальное количество кластеров?
4. Как оценить качество кластеризации и какие метрики используются для этого?

5. Что такое иерархическая кластеризация и как она отличается от других методов?
6. Что такое k-средних и как он используется в кластеризации?
7. Какие алгоритмы кластеризации используются для больших наборов данных?
8. Какие проблемы могут возникнуть при кластеризации и как их решить?
9. Как кластеризация используется в машинном обучении?
10. Как кластеризация может быть применена на практике в различных областях?

## Тема 4.2. Ассоциативные правила

### *Перечень изучаемых вопросов:*

Ассоциативные правила. Поддержка и уверенность. Правила ассоциации первого порядка. Правила ассоциации высоких порядков. Методы извлечения правил ассоциации. Алгоритм Apriori. Алгоритм FP-growth. Параллельные алгоритмы извлечения правил ассоциации. Построение дерева принятия решений на основе правил ассоциации. Использование правил ассоциации в рекомендательных системах. Анализ последовательностей событий на основе правил ассоциации. Оценка качества правил ассоциации.

### *Методические указания к изучению:*

Ассоциативные правила являются одним из ключевых методов анализа данных, широко используемых в области науки о данных и маркетинга. Эти правила представляют собой наборы правил вида «если А, то В», которые помогают выявлять связи между различными элементами набора данных. Одним из основных применений ассоциативных правил является анализ корзины покупок, который позволяет выявлять комбинации товаров, часто покупаемые вместе. Это полезно для магазинов, которые могут использовать такую информацию для оптимизации расположения товаров и увеличения продаж. Одним из основных преимуществ ассоциативных правил является их простота и понятность. Ассоциативные правила легко понимать и использовать, даже для людей без специальной подготовки в области науки о данных. Кроме того, ассоциативные правила могут использоваться для выявления скрытых закономерностей в данных, которые могут быть незаметными при первом взгляде. В этом разделе студенты рассмотрят основы ассоциативных правил, включая методы, такие как Apriori и FP-деревья. Студенты также рассмотрят применения ассоциативных правил в различных областях, включая маркетинг, медицину и финансы.

### *Литература:*

В предлагаемой литературе [1, 2, 8, 9] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

*Контрольные вопросы:*

1. Что такое ассоциативные правила и как они используются в анализе данных?
2. Что такое поддержка и уверенность в контексте ассоциативных правил?
3. Чем отличаются правила ассоциации первого порядка от правил ассоциации высокого порядка?
4. Какие методы извлечения правил ассоциации существуют и как они работают?
5. Как работает алгоритм Apriori в извлечении правил ассоциации?
6. Как работает алгоритм FP-growth в извлечении правил ассоциации?
7. Какие техники оптимизации времени и памяти используются в алгоритмах извлечения правил ассоциации?
8. Как оценивается качество правил ассоциации и какие метрики используются для этого?
9. Как правила ассоциации могут быть применены в различных областях, таких как маркетинг, медицина, финансы и экономика?
10. Какие проблемы могут возникнуть при использовании правил ассоциации и как их решить?

### **3.5. Раздел 5. Искусственные нейронные сети**

*Перечень изучаемых вопросов:*

Искусственные нейронные сети. Математические модели. Распознавание образов. Классификация данных. Прогнозирование. Обучение на основе данных. Многослойный перцептрон. Сверточные нейронные сети. Компьютерное зрение. Обработка естественного языка. Автономная навигация. Архитектура нейронных сетей. Интерпретируемость результатов. Шум в данных. Распознавание речи. Обработка изображений.

*Методические указания к изучению:*

Искусственные нейронные сети – это одна областей искусственного интеллекта. Они представляют собой математические модели, которые имитируют работу мозга и могут использоваться для решения широкого круга задач, таких как распознавание образов, классификация данных, прогнозирование и управление. Одним из основных преимуществ искусственных нейронных сетей является их способность к обучению на основе данных. Нейронные сети способны обучаться на больших объемах данных и

находить скрытые взаимосвязи между ними, что может быть невозможно для человека. Это позволяет использовать искусственные нейронные сети для решения сложных задач, таких как распознавание речи и обработка изображений. В данной разделе студенты рассмотрят основы искусственных нейронных сетей, включая методы, такие как перцептрон, многослойный перцептрон и сверточные нейронные сети. Студенты также рассмотрят применения искусственных нейронных сетей в различных областях, включая компьютерное зрение, обработку естественного языка и автономную навигацию.

#### *Литература:*

В предлагаемой литературе [1, 2, 4, 5, 6, 8] студенту необходимо для освоения темы изучить относящиеся к данной теме главы и разделы.

#### *Контрольные вопросы:*

1. Что такое искусственные нейронные сети и для чего они используются?
2. Какой принцип работы лежит в основе искусственных нейронных сетей?
3. Какие типы искусственных нейронных сетей существуют и в чем их отличия?
4. Как выбрать подходящую архитектуру нейронной сети для конкретной задачи?
5. Как происходит обучение искусственной нейронной сети и какие существуют методы обучения?
6. Как классифицировать задачи, решаемые с помощью искусственных нейронных сетей?
7. Что такое глубокое обучение и в чем его отличие от классического обучения нейронных сетей?
8. Какие существуют алгоритмы глубокого обучения и как их применять?

## 4. ТРЕБОВАНИЯ К АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

### 4.1. Текущая аттестация

Преподаватель вправе выбрать методику оценивания знаний студентов: традиционная зачетно-экзаменационная либо балльно-рейтинговая. При выборе методики должно учитываться мнение студентов. В случае если преподаватель выбрал балльно-рейтинговую систему, отдельные студенты вправе просить оценить их знания в рамках традиционной системы.

В рамках балльно-рейтинговой системы зачет выставляется по баллам, набранным: за выполнение текущих тестов, за качество выполнения и защиту лабораторных работ.

Тест считается пройденным успешно, если получено от 65 до 100 % верных ответов.

В таблице представлены виды деятельности и соотношение трудоемкости (для дневной формы обучения).

Вид деятельности	Доля	Кол-во ед.	Макс. балл за ед.	Всего
<b>Обязательные виды деятельности</b>				
1 семестр				
Посещаемость занятий	25 %	N1	=150/N1	150
Выполнение лаб. работ (защита)	60 %	7	60	360
Текущее тестирование	15 %	3	30	90
Итого:	100 %			600
<b>Дополнительные задания (по выбору студента в каждом семестре)</b>				
Подготовка реферата (видеодоклада)	20 %		120	120
Выполнение задания в рамках НИРС	40 %		240	240

### 4.2. Условие получения положительной оценки

Завершающим этапом изучения дисциплины является промежуточная аттестация в форме зачета. Оценка «зачтено» выставляется в случае, если студентом набрано не менее 600 баллов – в случае применения БРС.

В случае традиционного подхода к оцениванию оценка «зачтено» выставляется, если:

- для очной формы обучения – выполнены и защищены все лабораторные работы, пройдено текущее тестирование;

- для заочной формы обучения – выполнены и защищены все лабораторные работы, пройдено текущее тестирование, выполнена и зачтена контрольная работа.

## 5. ЗАКЛЮЧЕНИЕ

В учебно-методическом пособии нашли отражение следующие характеристики дисциплины «Технологии Data Mining»:

- 1) тематический план;
- 2) содержание дисциплины и указания к ее изучению, включающие по каждой теме: перечень изучаемых вопросов, методические указания к изучению темы, ссылки на литературу, контрольные вопросы;
- 3) требования к аттестации по дисциплине: содержание текущей аттестации, условия получения положительной оценки по промежуточной аттестации.

## 6. ЛИТЕРАТУРА

1. Цехановский, В. В. Технология интеллектуального анализа данных в процессах и системах / В. В. Цехановский, В. Д. Чертовской. – Санкт-Петербург : Лань, 2023. – 168 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/302753> (дата обращения: 23.02.2023). – ISBN 978-5-507-45404-4. – Текст : электронный.

2. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. – 2-е изд., стер. – Санкт-Петербург : Лань, 2022. – 212 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/206711> (дата обращения: 23.02.2023). – ISBN 978-5-8114-4493-9. – Текст : электронный.

3. Лаврищева, Е. М. Парадигмы моделирования и программирования задач предметных областей знаний / Е. М. Лаврищева, И. Б. Петров, А. К. Петренко ; под ред. А. И. Аветисян, О. Е. Баксанского, М. М. Горбунов-Посадова ; Институт системного программирования им. Иванникова [и др.]. – Москва ; Берлин : Директ-Медиа, 2021. – 504 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=602516> (дата обращения: 23.02.2023). – ISBN 978-5-4499-1889-5. – Текст : электронный.

4. Карпова, Т. С. Базы данных: модели, разработка, реализация / Т. С. Карпова. – 2-е изд., исправ. – Москва : Национальный Открытый Университет «ИНТУИТ», 2016. – 241 с.

5. Макарова, Н. В. Информатика : учеб. для вузов / Н. В.Макарова, В.Б.Волков. – СПб. : Питер, 2011. – 517 с.

6. Симонович, С. В. Информатика: базовый курс : учебное пособие / под ред. С. В. Симонович. – СПб. : Питер, 2008. – 639 с.

7. Грошев, А. С. Информатика : учеб. для вузов / А. С. Грошев, П. В. Закляков. – М. : ДМК Пресс, 2019. – 674 с.

8. Воробович, Н. П. Структуры данных: учебное пособие / Н. П. Воробович. – Министерство образования Российской Федерации, СибГТУ. – Красноярск : СибГТУ, 2002. – 119 с.

Локальный электронный методический материал

Михаил Андреевич Романов,  
Александр Борисович Тристанов

Технологии Data Mining

*Редактор М. А. Дмитриева*

Уч.-изд. л. 0,8. Печ. л. 1,1.

Издательство федерального государственного бюджетного  
образовательного учреждения высшего образования  
«Калининградский государственный технический университет».  
236022, Калининград, Советский проспект, 1.