

## **Тема 1. Введение в анализ данных**

1. Аналитический подход к моделированию. Движение от модели к результату. Информационный подход к моделированию. Построение модели от данных.

2. Информационный подход к моделированию. Построение модели от данных. Этапы моделирования. Тиражирование знаний. Процесс построения модели.

3. Процесс построения модели. Формулирование цели моделирования. Подготовка и сбор данных. Поиск модели.

4. Структурированные данные. Степени структурированности. Формализация данных. Принципы сбора данных.

5. Технология извлечения знаний из баз данных (Knowledge Discovery in Databases – KDD). Последовательность шагов, выполняемых в процессе KDD.

6. Технология добыча данных (Data Mining). Классификация задач Data Mining. Машинное обучение.

## **Тема 2. Интеллектуальный анализ данных: задача ассоциации**

1. Аффинитивный анализ. Примеры ассоциативных правил. Транзакция. Предметный набор. Ассоциативное правило.

2. Ассоциативное правило. Поддержка ассоциативного правила. Достоверность ассоциативного правила. Примеры.

3. Ассоциативное правило. Значимость ассоциативных правил. Лифт. Левередж. Примеры.

4. Поиск ассоциативных правил. Алгоритм Apriori. Частый предметный набор. Пример.

## **Тема 3. Интеллектуальный анализ данных: задача кластеризации**

1. Кластеризация. Постановка задачи кластеризации. Задачи, решаемые кластеризацией.

2. Кластеризация. Алгоритм k-means. Шаги алгоритма k-means. Евклидово расстояние. Расстояние Манхеттена.
3. Кластеризация. Алгоритм k-means. Евклидово расстояние. Расстояние Манхеттена. Критерий сходимости. Недостатки алгоритма k-means.
4. Алгоритм G-means. Недостатки алгоритма k-means.
5. Кластеризация. Проблемы алгоритмов кластеризации. Выбор числа кластеров.

#### **Тема 4. Интеллектуальный анализ данных: байесовский классификатор.**

1. Простой байесовский классификатор. Теоретические сведения.
2. Байесовские сети. Байесовские сети с двумя и тремя переменными. Примеры.
3. Байесовские сети. Последовательная связь. Примеры.
4. Байесовские сети. Расходящаяся связь. Примеры.
5. Байесовские сети. Сходящаяся связь. Примеры.

#### **Тема 5. Интеллектуальный анализ данных: дерево решений**

1. Дерево решений. Понятия, встречающиеся в теории деревьев решений. Разделение на классы. Пример.
2. Дерево решений. Условия построения дерева решений. Факторы популярности дерева решений.
3. Дерево решений. Структура дерева решений. Выбор атрибута разбиения в узле.
4. Дерево решений. Процесс рекурсивного разбиения (разделяй и властвуй). Варианты разбиения.
5. Дерево решений. Полное дерево решений. Меры эффективности деревьев решений.
6. Дерево решений. Критерии выбора наилучших атрибутов ветвления. Индекс Джини.

7. Дерево решений. Критерии выбора наилучших атрибутов ветвления. Уменьшение энтропии, или прирост информации.

### **Тема 6. Интеллектуальный анализ данных: линейная регрессия**

1. Регрессия. Линейная и логистическая регрессия. Примеры.
2. Простая линейная регрессия. Линия регрессии. Уравнение регрессии. Коэффициенты регрессии.
3. Регрессия. Оценка соответствия простой линейной регрессии реальным данным. Стандартная ошибка.
4. Регрессия. Изменчивость выходной переменной. Меры, характеризующие поведение выходной переменной. Примеры.
5. Регрессия. Коэффициент детерминации. Коэффициент корреляции. Примеры.

### **Тема 7. Интеллектуальный анализ данных: логическая регрессия**

1. Основы логистической регрессии. Простой пример логистической регрессии.
2. Основы логистической регрессии. Построение линии логистической регрессии.
3. Основы логистической регрессии. Оценки максимального правдоподобия.
4. Основы логистической регрессии. Значимость входных переменных.
5. Основы логистической регрессии. Использование логистической регрессии для решения задач классификации.
6. Интерпретация модели логистической регрессии. Шансы и отношение шансов.

### **Тема 8. Анализ и прогнозирование временных рядов**

1. Прогнозирование. Временной ряд и его компоненты. Непрерывный и дискретный временные ряды. Цели и задачи анализа временных рядов.

2. Прогнозирование. Детерминированная и случайная составляющая временного ряда. Модели временных рядов.

3. Прогнозирование. Компоненты временного ряда. Тренд. Сезонная и циклические компоненты.

4. Прогнозирование. Исследование временных рядов и автокорреляция. Коэффициент автокорреляции. Примеры.

5. Модели прогнозирования. Обобщенная модель прогноза. Метод скользящего окна.

6. Методы прогнозирования: формализованные, эвристические и комплексные.

7. Модели прогнозирования. «Наивная» модель прогнозирования. Экстраполяция.

8. Модели прогнозирования. Прогнозирование методом среднего и скользящего среднего.

9. Модели прогнозирования. Регрессионные модели. Метод декомпозиции временного ряда.

10. Ансамбли моделей. Комбинирование решений. Виды ансамблей.

11. Ансамбли моделей. Бэггинг - основная идея. Алгоритм и схема процедуры бэггинга.

12. Ансамбли моделей. Бустинг – основная идея. Процедура бустинга. Отличие от бэггинга.