

## **Анализ данных - поиск закономерностей и построение моделей прогноза**

**Код:** REQ-026

**Длительность:** 16 ч.

### **Описание:**

На курсе рассматриваются основные типы моделей, используемые для поиска закономерностей в данных: регрессионный анализ, кластеризация данных, построение простых и обобщенных деревьев решений, сокращение данных – метод главных компонент.

Эти методы, входящие в научную дисциплину «Распознавания образов» (Pattern Recognition), являются основой такой быстро развивающейся дисциплины как Business Intelligence (BI) и широко используются в бизнес-аналитике.

Кратко обсуждаются идеи и методы нейронных сетей и машины поддерживающих векторов (SVM – Support Vector Machine), а также методы bootstrap-построения оценок при недостаточном числе исходных данных. Обсуждаются основные понятия «нечеткого» (fuzzy) анализа данных.

Предлагаемые алгоритмы могут быть использованы в базах данных для восстановления (предсказания) пропущенных значений.

### **Цели:**

Целью курса является знакомство слушателей с постановками задач поиска зависимостей и распознавания образов, описание математических моделей и разбор пошаговых действий (алгоритма) их решения. Здесь описаны процедуры проверки прогностической устойчивости моделей и правила определений области допустимых значений данных, поступающих для прогнозирования. По окончании курса слушатели научатся понимать применимость основных методов анализа данных, например, при восстановлении пропущенных значений в БД и в задачах классификации и распознавания образов.

### **Разбираемые темы:**

- Введение в анализ данных и распознавание образов;
- Первичное преобразование данных, поиск выбросов;
- Регрессионный анализ. Скользящий контроль;
- Деревья решений. Простая и обобщенная формы;
- Кластеры и их поиск;

- Кластер как связная компонента графа;
  - Построение минимального покрывающего дерева;
  - Метод К средних. Простая и обобщенная версии;
  - Иерархический кластер анализ. Дендрограммы;
- 
- Метод главных компонент. Факторы и их поиск;
  - Продвинутое методы анализа;
- 
- Нейронные сети;
  - SVM и поддерживающие вектора. Кernels функции;
  - Эволюционные алгоритмы – МГУА, генетические;
  - Метод Bootstrap;
  - Семейства прогнозирующих алгоритмов;
  - «Нечеткие» классификаторы.

**Целевая аудитория:**

Аналитики, бизнес-аналитики, разработчики, руководители групп, нуждающиеся в кратком и доступном изложении методов анализа данных.

**Предварительная подготовка - общее:**

Требуется знание терминов линейной алгебры и правил работы с матрицами и векторами на уровне 1-2 курса технического Вуза.

**Примечание:**

Материалы курса представлены на английском языке.