

## **Б1.В.ДВ.07.02 Технологии машинного обучения и анализа больших данных**

### **Цели дисциплины**

Целями освоения дисциплины Технологии машинного обучения и анализа больших данных является формирование профессиональных компетенций будущих специалистов в области теоретических основ методов анализа больших данных и машинного обучения. Опираясь на знания, полученные при изучении курсов высшей математики, программирования и численных методов рассмотреть основы решения прикладных задач математики методами анализа больших данных и машинного обучения.

### **Задачи дисциплины**

Основными задачами изучения дисциплины являются:

- Формирование знаний в области анализа больших данных и машинного обучения, методов решения экспериментальных и теоретических задач;
- Формирование практических навыков применения математического аппарата и систем анализа больших данных для решения задач машинного обучения;
- Формирование навыков владения интерпретирования, представления и применения результатов, полученных с использованием анализа больших данных и машинного обучения для своей предметной области.

### **Формируемые компетенции и индикаторы их достижения по дисциплине**

<b>Код компетенции</b>	<b>Содержание компетенции</b>	<b>Код и наименование индикатора достижения компетенции</b>
ПКС-2.	Способен проводить формализацию предметной области с целью создания информационной системы	ПКС-2.1 - Знает требования к компьютерному программному обеспечению; виды технической спецификации на программные компоненты и их взаимодействие; методы проектирование компьютерного программного обеспечения ПКС-2.2 – Умеет применять требования к компьютерному программному обеспечению; разрабатывать технические спецификации на программные компоненты и их взаимодействие; применять методы проектирования компьютерного программного обеспечения; ПКС-2.3 – Владеет методами разработки требований к компьютерному программному обеспечению, технических спецификаций на программные компоненты, методами проектирования компьютерного программного обеспечения.

### **Содержание разделов дисциплины**

#### **Тема 1 Введение в машинное обучение и обработку данных. Программный инструментарий Data Mining и Machine Learning.**

Введение в машинное обучение и обработку данных. Постановка основных классов задач в машинном обучении. Обучение с учителем; регрессия и классификация; обучение без учителя; кластеризация, снижение размерности; рекомендательные системы; обработка текстов; обработка изображений; обучение с подкреплением. Примеры задач. Программный инструментарий Data Mining и Machine Learning

## **Тема 2 Основной аппарат комбинаторики и мат. статистики. Регрессионный анализ и сжатие данных.**

Основной аппарат комбинаторики и мат. статистики. Основные понятия математической статистики: статистические оценки, их свойства, проверка гипотез. Регрессионный анализ и сжатие данных. Задача регрессии. Минимизация квадрата отклонения. Регрессионная функция: условное мат.ожидание. Линейная регрессия и метод k ближайших соседей. Переобучение и недообучение. Разложение ошибки на шум, смещение и разброс.

## **Тема 3 Детектирование выбросов и аномалий. Очистка данных и технологии регуляризации**

Детектирование выбросов и аномалий. Что такое выбросы, типы выбросов. Методы обнаружения выбросов. Поиск аномалий. Цензурирование выборки. Отсев объектов-выбросов, удаление выбросов. Очистка данных и технологии регуляризации. Основные виды регуляризации. Метод редукции размерности. Методы отбора признаков.

## **Тема 4 Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы**

Технологии кластеризации и классификации. K-means. EM-алгоритм. Другие методы кластеризации. Задачи классификации. Байесовский классификатор. Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия. Нейронные сети: общая архитектура. Многослойные сети. Обратное распространение ошибки. Стохастический градиентный спуск. Генетические алгоритмы.

## **Тема 5. Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети**

Извлечение признаков / выделение особенностей (Feature detection). Преобразования признаков. Нормализация данных. Методы нормализации данных. Нормализация по методу минимакса. Нормализация по Z-показателю. Десятичное масштабирование. Нечеткие множества. Байесовы сети. Задачи байесовского вывода. Методика построения нечеткой байесовой сети