

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ
Б1.В.04 «Машинное обучение и искусственный интеллект»*наименование дисциплин (модуля)/практики***Автор:** ПАО «Сбербанк», руководитель направления по анализу данных, Заворина Е.И.**Код и наименование направления подготовки, программа бакалавриата:**

38.03.06 Торговое дело, «Цифровизация бизнеса и электронная торговля»

Квалификация (степень) выпускника: Бакалавр**Форма обучения:** очная

Цель освоения дисциплины: сформировать компетенции для проведения аналитического исследования с применением технологий больших данных с использованием существующей в организации методологической и технологической инфраструктуры (ПКс ОС-6)

План курса:**Тема 1. Введение в машинное обучение**

История анализа данных. Постановки задач в машинном обучении: классификация, регрессия, ранжирование, кластеризация, латентные модели. Примеры задач. Виды данных: структурированные таблицы, тексты, изображения, звук, логи. Признаки.

Тема 2. Линейные методы регрессии

Аналитическое и численное решение задачи МНК. Градиентный спуск, методы оценивания градиента. Функции потерь. Квантильная регрессия (постановка задачи и примеры использования). Методы оценивания обобщающей способности, кросс-валидация. Метрики качества регрессии. Оценивание качества скользящим окном.

Тема 3. Линейные методы классификации

Аппроксимация эмпирического риска. Персептрон. Метод опорных векторов, его двойственная задача (без ядер). Задача оценивания вероятностей, логистическая регрессия. Идея калибровки вероятностей. Оптимизация второго порядка (идея и предпосылки для использования). Обобщённые линейные модели. Метрики качества в задачах классификации. Multiclass- и multilabel-классификация. Особенности многоклассовых задач. Метрики качества. Методы решения multilabel-задач, основанные на матричных разложениях

Тема 4. Особенности работы с реальными данными

Пропуски в данных. Предобработка признаков. Чистка данных. Категориальные признаки: кодирование, хэширование, счётчики. Работа с текстами. Разреженные признаки: векторизация, хэширование, TF-IDF. Косинусная метрика.

Тема 5. Композиции алгоритмов

Общая идея bias-variance decomposition. Бэггинг и метод случайных подпространств. Случайные леса и extra random trees. Бустинг. Градиентный бустинг над решающими деревьями. Модель xgboost.

Тема 6. Введение в искусственный интеллект

Основные направления исследований в области искусственного интеллекта. Нейросетевой подход к созданию интеллектуальных систем. Понятие экспертной системы (ЭС).

Тема 7. Базы знаний

База знаний – основная компонента экспертной системы. Отличия знаний от данных, базы знаний от базы данных. Архитектура ЭС. Отличия ЭС от традиционных программных систем. Основные типы решаемых задач и области применения ЭС.

Тема 8. Инженерия знаний

Технологии инженерии знаний. Классификация методов извлечения знаний. Примеры систем приобретения знаний. Представление нечетких знаний. Вывод в условиях неопределенности

Формы текущего контроля и промежуточной аттестации:

Формы текущего контроля успеваемости: эссе.

Формы промежуточной аттестации, отражающие формирование компетенции на уровне данной дисциплины: **зачет**.

**Компетенция ПКс ОС-6 отражает формирование:
на уровне знаний:**

Методы и инструментальные средства управления аналитическими проектами по исследованию больших данных

Содержание и последовательность выполнения этапов аналитического проекта по исследованию больших данных

Типы анализа больших данных, виды аналитики

Теоретические и прикладные основы анализа больших данных

Современные методы и инструментальные средства анализа больших данных

на уровне умений:

Представлять содержание и результаты работ по анализу больших данных

Вести протоколы мероприятий по анализу больших данных

Планировать аналитические работы с использованием технологий больших данных

Проводить аналитические работы с использованием технологий больших данных

Осуществлять интеграцию и преобразование данных в ходе работ по анализу больших данных

на уровне навыков:

Разработка, обсуждение и утверждение содержания аналитических работ с использованием технологий больших данных

Определение состава группы для проведения анализа больших данных

Определение необходимых ресурсов для проведения аналитических работ

Разработка, обсуждение и утверждение плана аналитических работ

Распределение ролей и состава аналитических работ между участниками группы для анализа больших данных

Основная литература:

1. Жданов А.А. Автономный искусственный интеллект / Жданов А.А. — Москва: Лаборатория знаний, 2020. — 360 с. — ISBN 978-5-00101-655-7. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/6506.html> (дата обращения: 08.04.2021). — Режим доступа: для авторизир. пользователей
2. Барский А.Б. Искусственный интеллект и логические нейронные сети: учебное пособие / Барский А.Б. — Санкт-Петербург: Интермедия, 2019. — 360 с. — ISBN 978-5-4383-0155-4. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/95270.html> (дата обращения: 08.04.2021). — Режим доступа: для авторизир. пользователей
3. Джеймс Баррат Последнее изобретение человечества: искусственный интеллект и конец эры Homo sapiens / Джеймс Баррат. — Москва: Альпина нон-фикшн, 2019. — 312 с. — ISBN 978-5-91671-436-4. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/86821.html> (дата обращения: 08.04.2021). — Режим доступа: для авторизир. пользователей