Методические материалы по практическим работам

Д.И. Каплун, А.С. Вознесенский

АНАЛИЗ ДАННЫХ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ

# Практические работы

## Общие положения и методика оценивания результатов

В процессе обучения по дисциплине «Анализ данных в искусственном интеллекте» студент обязан выполнить 8 практических работ в формате онлайн тестирования на сайте mathworks.com (доступ через личный аккаунт). Под выполнением практических работ подразумевается подготовка к работе, проведение экспериментальных исследований, подготовка отчета и его защита. Выполнение практических работ студентами осуществляется индивидуально. Оформление отчета студентами осуществляется индивидуально в соответствии с принятыми в СПбГЭТУ правилами оформления студенческих работ. Отчет оформляется после выполнения экспериментальных исследований и представляется преподавателю на проверку. После проверки отчет либо возвращается (при наличии замечаний) на доработку, либо подписывается к защите. Практические работы защищаются студентами индивидуально. Каждый студент получает вопрос по теоретической и практической части, после чего ему предоставляется время для подготовки ответа. На защите практической работы студент должен показать: понимание и умение объяснять особенности применяемых методов и алгоритмов, возможные области их применения и т.д., прогнозировать реакции исследуемого объекта на различные воздействия, навыки и умения, приобретенные при выполнении практических работы. При обсуждении ответа преподаватель может задать несколько уточняющих вопросов. В случае если студент демонстрирует достаточное знание вопроса, работа считается зачтенной. Текущий контроль включает в себя выполнение, сдачу в срок отчетов и их защиту по всем практическим работам, по результатам которой студент получает допуск к экзамену (итоговое тестирование). Критерием оценки работы на коллоквиумах является оценка, выставляемая по 5-ти балльной шкале в соответствии со следующими критериями: оценка в 5 баллов выставляется при отличном выполнении задания, то есть при наличии полных (с детальными пояснениями и выкладками), оригинальных и правильных решений задач, дополненных при необходимости документами, полученными в результате реализации (проверки) решения, верных ответов и высококачественного оформления работы. Оценка в 4 балла выставляется при правильном выполнении задания, то есть при наличии полных (с пояснениями и выкладками), оригинальных и правильных решений задач, дополненных при необходимости документами, полученными в результате реализации (проверки) решения, верных ответов. Оценка в 3 балла выставляется при наличии отдельных неточностей в ответах (включая грамматические ошибки) или неточностях в решении задач непринципиального характера (описки и случайные ошибки арифметического характера). Оценка в 2 и ниже баллов выставляется в случаях, когда в ответах и в решениях задач имеются неточности и ошибки, свидетельствующие о недостаточном понимании вопросов и требующие дополнительного обращения к тематическим материалам.

Контроль самостоятельной работы студентов осуществляется на практических занятиях студентов по методикам, описанным выше.

## Содержание отчетов

* Титульный лист
* Формулировка задания
* Краткое описание выполнения основных этапов работы.
* Результаты работы
* Выводы по работе

## Практические работы

### Искусственный интеллект и машинное обучение

**Цель работы**: ознакомить студентов с основными терминами в области искусственного интеллекта и машинного обучения.

**Задание:** тест [MATLAB Onramp](https://matlabacademy.mathworks.com/details/matlab-onramp/gettingstarted)

**Контрольные вопросы:** Искусственный интеллект. Машинное обучение. Общая постановка задачи обучения по прецедентам. Типология обучения задач по прецедентам. Бустинг. Бэггинг. Области применения ИИ.

### Сбор, хранение и подготовка данных. Дополнение данных, заполнение (удаление) пропущенных значений

**Цель работы**: ознакомить студентов с основными способами получения и подготовки данных.

**Задание:** тест Signal Processing Onramp

**Контрольные вопросы:** Как подготовить данные к моделированию: 5 операций Data Preparation. Почему нужно готовить данные к моделированию. Как Data Scientist готовит данные: методы и средства Data Preparation. Отберем то, что нужно Data Mining: как сформировать датасет для машинного обучения. Что такое датасет для Data Mining и из чего он состоит. Где искать датасеты. Каким бывает dataset: типы выборок. Как сформировать выборку для Data Mining. Зачем нужна очистка данных для Data Mining: 10 главных проблем подготовки датасета и способы их решения. Что такое очистка данных для Data Mining. От чего надо чистить большие данные. Как Data Scientist чистит данные: детали фазы Data Preparation Это не баг, а фича: генерация признаков для Data Mining. Признаки для Data Mining: определение и виды. Как проходит генерация признаков: 3 задачи этого этапа Data Mining и способы их решения. Data Preparation: полет нормальный – что такое нормализация данных и зачем она нужна.

### Подготовка данных. Выявление аномалий в сырых данных, очистка данных от шума

**Цель работы**: ознакомить студентов с основными методами и алгоритмами исключения аномалий в данных и очистки данных от шума.

**Задание:** тест Image Processing Onramp

**Контрольные вопросы:**

Сырые данные. Выбросы. Причины выбросов. Определение выбросов. Критериии выбросов I. Гистограмма. Ящик с усами. Критерий Шовене. Критерий Граббса. Критерий Пирса. Критерий Диксона. Правило 3 сигм. Критерии выбросов II. Статистические тесты. Модельные тесты. Итерационные методы. Метрические методы. Методы подмены задачи. Методы машинного обучения. Ансамбли алгоритмов. ROC анализ. Алгоритмы очистки сигналов от шума. Преобразование Фурье. Вейвлет преобразование. Фильтр Винера. Алгоритмы Empirical Mode Decomposition и Intrinsic Time Scale Decomposition. Алгоритмы очистки изображений от шума. Типовые шумы. Виды шумов: аналоговый, цифровой. Методы шумоподавления: пространственные, временные, пространственно временные. Качество шумоподавления. Метрики качества: PSNR и SSIM. Усредняющий фильтр. Медианный фильтр. Гауссовский фильтр. Билатеральный фильтр. Алгоритм NLM.

### Классификация

**Цель работы**: ознакомить студентов с понятием задачи классификации и основными алгоритмами классификации.

**Задание:** тест Optimization Onramp

**Контрольные вопросы:** Задача классификации. Типы входных данных. Типы классов. Области применения. Формальная постановка задачи. Алгоритмы классификации. Метод ближайших соседей. Байесовский классификатор. Деревья решений.

### Кластеризация

**Цель работы**: ознакомить студентов с понятием задачи кластеризации и основными алгоритмами кластеризации.

**Задание:** тест [Wireless Communications Onramp](https://matlabacademy.mathworks.com/details/wireless-communications-onramp/wireless)

**Контрольные вопросы:** Задача кластеризации. Типы входных данных. Цели кластеризации. Функции расстояния. Формальная постановка задачи. Алгоритмы кластеризации. Метод k-средних. Метод c-средних. Генетический алгоритм.

### Нейронные сети

**Цель работы**: ознакомить студентов с понятием нейронной сети и терминологией.

**Задание:** тест [Machine Learning Onramp](https://matlabacademy.mathworks.com/details/machine-learning-onramp/machinelearning)

**Контрольные вопросы:** Что такое нейронная сеть (НС)? Хронология. Классификация НС. Этапы решения задачи. Области применения НС. Структура нейрона. Синапс. Принципы работы НС. Функция активации. Обучение, валидация, тестирование. Итерация. Эпоха. Ошибка. Нейрон смещения. Обучение НС. Градиентный спуск. Метод обратного распространения (МОР). Гиперпараметры. Cходимость. Переобучение.

### Сверточные нейронные сети

**Цель работы**: ознакомить студентов с понятием сверточной нейронной сети и терминологией.

**Задание:** тест [Deep Learning Onramp](https://matlabacademy.mathworks.com/details/deep-learning-onramp/deeplearning)

**Контрольные вопросы:** Сверточные нейронные сети (СНС). Архитектура СНС. Слой свертки. Слой активации. Пулинг или субдискретизирующий слой. Полносвязная НС. Обучение СНС. Преимущества СНС. Недостатки СНС.

### Использование современной вычислительной базы в контексте решения задач Data Science (CUDA, GPU, FPGA, SoC). Современные программно-аппаратные средства (MATLAB, Python, R, GPU)

**Цель работы**: ознакомить студентов с использованием современной программно-аппаратной базы в контексте решения задач Data Science.

**Задание:** тест [Simulink Onramp](https://matlabacademy.mathworks.com/details/simulink-onramp/simulink)

**Контрольные вопросы:** Цифровые процессоры обработки сигналов (ЦПОС). Графические процессоры (GPU). Программируемые логические интегральные схемы (ПЛИС). Технология CUDA.Системы на кристалле (SoC). Языки математических и научных расчетов: MATLAB, Python, R.