

Департамент информатизации Тюменской области
Государственное автономное учреждение дополнительного образования
Тюменской области
«Региональный информационно-образовательный центр»

СОГЛАСОВАНО

Исполнительный директор
ГАУ ТО «Западно-Сибирский
инновационный центр»



А.В. Сакевич

« 09 » июня 2020 г.

УТВЕРЖДАЮ

Директор
ГАУ ДО ТО «РИО-Центр»



Т.А. Беляева

« 09 » июня 2020 г.

УЧЕБНАЯ ПРОГРАММА

«Подготовка и поиск обучающих данных для нейросетевых моделей»

Трудоемкость программы – 56 академических часов

Форма обучения – очная, заочная с применением дистанционных образовательных технологий

Режим занятий – 6 академических часов в неделю

Начальные навыки: Базовые навыки работы на персональном компьютере

Цель обучения: сформировать у слушателей комплекс знаний, умений и практических навыков сбора и подготовки данных для обучения нейронных сетей.

Компетенции на выходе:

Слушатели обладают всеми необходимыми знаниями, умениями и навыками программирования, а также навыками сбора и подготовки данных для обучения нейронных сетей.

Знания:

- знать место языка Python среди языков программирования высокого уровня,
- знать особенности структуры программы, представленной на языке Python,
- знать возможности и ограничения использования готовых модулей,
- знать что такое операция, операнд и их характеристики,
- знать принципиальные отличия величин структурированных и не структурированных,
- знать математические функции, входящие в Python,
- знать основные операторы языка Python, их синтаксис,
- знать правила описания процедур и функций в Python и построение вызова процедуры,
- знать принципиальные отличия между формальными, локальными и глобальными переменными,
- знать область действия описаний в процедурах,
- знать, как с помощью Списков определять в программе тип «массив», «матрица»
- знать свойства данных типа «массив», «матрица»

Умения:

- уметь записывать примеры арифметических и логических выражений всех атрибутов, которые могут в них входить;
- уметь разрабатывать программы обработки числовой и символьной информации;
- уметь разрабатывать программы (линейные, разветвляющиеся и с циклами);
- уметь воспроизводить алгоритмы сортировки массивов и матриц, поиска в упорядоченном массиве, распространять эти алгоритмы на сортировку и поиск в нечисловых массивах;
- уметь читать и записывать текстовые файлы в заданном формате;
- уметь собирать, анализировать информацию и строить визуальные отчеты;

- уметь создавать, обучать и применять искусственные нейронные сети;
- уметь подготавливать датасеты для обучения нейронных сетей;
- уметь оценивать качество нейросетевых моделей.

Приёмы работы:

- с модулями, входящими в состав среды Python;
- с величиной, ее характеристиками;
- с такими структурами данных, как число, текст, кортеж, список, словарь;
- с составом арифметического выражения;
- с логическими выражениями и входящих в них операндах, операциях и функциях;
- с процессами исполнения операторов;
- с основными приемами формирования процедуры и функции;
- с основными алгоритмическими задачами в среде Python.

Программой обучения предусмотрены лекционные занятия и практические занятия. Под лекционными занятиями подразумеваются аудиторные занятия, либо занятия с применением средств видео-конференц-связи с выдачей материала преподавателем. Под практическими занятиями подразумевается самостоятельное выполнение слушателями практических заданий в учебной аудитории либо с применением средств видео-конференц-связи в присутствии либо под контролем преподавателя.

УЧЕБНЫЙ ПЛАН

№ п/п	Наименование тем	Всего часов	в том числе:		Формы контроля
			Лекционные занятия	Практические занятия	
1.	Основные понятия теории вероятностей, теоретической и прикладной статистики	8	4	4	Лабораторная работа
2.	Основы программирования для решения задач разработки нейросетевых архитектур	18	8	10	Лабораторная работа
3.	Подготовка и поиск обучающих данных для нейросетевых моделей	30	9	21	Лабораторная работа
Итого часов:		56	21	35	

УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН

№ п/п	Наименование тем	Всего часов	в том числе:		Формы контроля
			Лекционные занятия	Практические занятия	
1.	1. Основные понятия теории вероятностей, теоретической и прикладной статистики	8	4	4	
1.1.	Основные понятия теории вероятностей	2	1	1	Лабораторная работа
1.2.	Статистические величины	2	1	1	
1.3.	Выборочное наблюдение	2	1	1	
1.4.	Базовые распределения, статистики и их свойства	2	1	1	
2.	Основы программирования для решения задач разработки нейросетевых архитектур	18	8	10	
2.1.	Философия языка Python. Базовый синтаксис. Переменные и выражения	2	1	1	Лабораторная работа
2.2.	Типы данных Python. Операции над ними	2	1	1	
2.3.	Структуры данных. Управляющие структуры	2	1	1	
2.4.	Обработка исключений. Функции	2	1	1	
2.5.	Структура проекта и модули	2	1	1	
2.6.	Строки и регулярные выражения	2	1	1	
2.7.	Работа с данными: текстовые файлы, csv, excel, pickle	1		1	
2.8.	Парсинг сайтов и основы HTML	5	2	3	

3.	Подготовка и поиск обучающих данных для нейросетевых моделей	30	9	21	
3.1.	Понятие о машинном обучении	2	2		Лабораторная работа
3.2.	Процесс создания, тренировки и тестирования искусственных нейронных сетей	2	2		
3.3.	Подготовка графических датасетов: поиск готовых графических датасетов, технология подготовки графического датасета «с нуля»	2	2		
3.4.	Подготовка текстовых датасетов: поиск готовых датасетов, технология подготовки текстового датасета, особенности языковых конструкций при подготовке текстовых датасетов	2	2		
3.5.	Практикум по работе над датасетами (разметка датасетов, подготовка датасетов работа по подготовке шаблонизированного текстового датасета в OCAS DeepData)	18		18	
3.6.	Оценка качества нейросетевых моделей. Основные метрики точности искусственной нейронной сети	4	1	3	
	Всего часов:	56	21	35	

СОДЕРЖАНИЕ ПРОГРАММЫ

Раздел 1. Основные понятия теории вероятностей, теоретической и прикладной статистики.

1.1. Основные понятия теории вероятностей

Знакомство с понятиями и законами теории вероятности и комбинаторики.

1.2. Статистические величины

Определение абсолютных и относительных случайных величин и их свойств.

1.3 Выборочное наблюдение

Требования к выборочным наблюдениям, принципы и способы организации выборки

1.4 Базовые распределения, статистики и их свойства

Основные распределения в статистике, их свойства и применение.

Раздел 2. Основы программирования для решения задач разработки нейросетевых архитектур.

2.1 Философия языка Python. Базовый синтаксис. Переменные и выражения

Python дзен, синтаксис языка, понятие переменной.

2.2 Типы данных Python. Операции над ними

Типы данных, их отличие, преобразование и возможные операции.

2.3 Структуры данных

Понятие списков, кортежей, множеств, словарей и массивов, их различия.

2.4 Управляющие структуры

Конструкции условия и циклы. Их синтаксис и применение.

2.5 Обработка исключений. Функции

Типы и обработка ошибок. Повторное использование кода, lambda-функции.

2.6 Структура проекта и модули

Понятие проекта и его правильная организация. Организация кода в модули. Импортирование кода.

2.7 Строки и регулярные выражения

Строки и работа с ними. Понятие регулярных выражений и их использование для работы с текстом.

2.8 Работа с данными: текстовые файлы, csv, excel, pickle

Работа с файлами. Чтение, запись и обработка информации в файлах.

2.9 Парсинг сайтов и основы HTML

Язык разметки HTML, структура WEB-страницы и основные теги. Сбор и обработка информации с HTML страниц.

Раздел 3. Подготовка и поиск обучающих данных для нейросетевых моделей

3.1 Понятие о машинном обучении

История, основные понятия и сферы применения машинного обучения.

3.2 Процесс создания, тренировки и тестирования искусственных нейронных сетей

Определение нейрона и нейронной сети. Как обучается нейронная сеть и типы обучения. Создание, обучение и тестирование собственных нейронных сетей.

3.3 Подготовка графических датасетов: поиск готовых графических датасетов, технология подготовки графического датасета «с нуля»

Поиск и сбор изображений для датасета. Основные ресурсы с готовыми датасетами.

3.4 Подготовка текстовых датасетов: поиск готовых датасетов, технология подготовки текстового датасета, особенности языковых конструкций при подготовке текстовых датасетов

Поиск и сбор текстов для датасета. Основные ресурсы с готовыми датасетами. Эмбединги.

3.5 Практикум по работе над датасетами (разметка датасетов, подготовка датасетов, работа по подготовке шаблонизированного текстового датасета в OCAS DeepData)

Подготовка датасета и его разметка для обучения нейронной сети.

Сервисы для подготовки датасетов. Подготовка текстового датасета для обучения нейронной сети. Работа в OCAS DeepData.

3.7 Оценка качества нейросетевых моделей. Основные метрики точности искусственной нейронной сети

Оценка и метрики обучения нейронной сети. Виды проблем обучения и их решение.