

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

САЕ “Институт человека цифровой эпохи”

УТВЕРЖДАЮ:  
Руководитель ОПОП:

 З.И. Резанова

« 31 » августа 20 22 г.

Рабочая программа дисциплины

**Машинное обучение**

по направлению подготовки

**45.04.03 Фундаментальная и прикладная лингвистика**

Направленность (профиль) подготовки :  
**Компьютерная и когнитивная лингвистика**

Форма обучения

**Очная**

Квалификация

**Магистр**

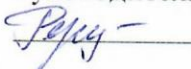
Год приема

**2022**


Код дисциплины в учебном плане: Б1.В.ДВ.1.1.3

СОГЛАСОВАНО:

Руководитель ОПОП

 З.И. Резанова

Председатель УМК

 Ю.А. Тихомирова

Томск – 2022

## **1. Цель и планируемые результаты освоения дисциплины**

Целью освоения дисциплины является формирование следующих компетенций:

ПК-1. Способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований.

ПК-3. Способность разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты электронных ресурсов и интеллектуальных электронных систем (лингвистические корпуса, словари, онтологии, базы данных).

ПК-4. Способность разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИПК-1.3. Последовательно реализует исследовательскую программу, получает новые научные результаты.

ИПК-3.1 Разрабатывает системы автоматической обработки звучащей речи и письменного текста на естественном языке.

ИПК-3.2 Разрабатывает лингвистические компоненты электронных ресурсов (лингвистические корпуса, словари).

ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

ИПК-4.3 Обеспечивает выполнение проекта в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта, в соответствии с установленными целями, сроками и затратами.

## **2. Задачи освоения дисциплины**

Студент, успешно освоивший дисциплину:

— демонстрирует уверенные и системные знания о принципах и особенностях программных решений и продуктов ведущих исследовательских команд в области машинного обучения.

— свободно и уверенно может создать проект интеллектуальной информационной системы или системы распознавания письменной речи.

— демонстрирует свободное владение навыками написания части программного кода системы анализа естественного языка.

## **3. Место дисциплины в структуре образовательной программы**

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор. Дисциплина входит в модуль Компьютерная лингвистика.

## **4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине**

Второй семестр, зачет с оценкой

## **5. Входные требования для освоения дисциплины**

Для успешного освоения дисциплины требуются результаты обучения по ранее освоенным дисциплинам.

## **6. Язык реализации**

Русский

## **7. Объем дисциплины**

Общая трудоемкость дисциплины составляет 4 з.е., 144 часов, из которых:

-лекции: 6 ч.

-практические занятия: 30 ч.

Объем самостоятельной работы студента определен учебным планом.

## **8. Содержание дисциплины, структурированное по темам**

Тема 1. Введение в машинное обучение.

Наблюдения, признаки, экземпляры, обучение с учителем, без учителя, смешанный типа обучения. Задачи регрессии и классификации.

Тема 2. Линейные модели.

Знакомство с библиотекой scikit-learn. Логистическая регрессия. Support Vector Machine. Функция потерь.

Тема 3. Метрические методы.

Продолжение работы с библиотекой scikit-learn. Метод k ближайших соседей. Дистанция в машинном обучении.

Тема 4. Решающие деревья.

Построение дерева решений. Условия остановки алгоритма.

Тема 5. Введение в NLP.

Предобработка данных, предобработка текста. Знакомство с библиотекой NLTK. Стоп-слова, лемматизация, токенизация, стемминг. Классификация текстов. Сентимент-анализ.

Тема 6. Нейронные сети.

Перцептрон. Многослойный перцептрон. Обратное распространение ошибки. Функции оптимизации и активации. Библиотеки Keras и Tensorflow.

Тема 7. Метрики.

Оценка работы системы машинного обучения. Матрица ошибок. Метрики в NLP.

## **9. Текущий контроль по дисциплине**

Текущий контроль по дисциплине проводится путем контроля посещаемости, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

## **10. Порядок проведения и критерии оценивания промежуточной аттестации**

Зачет с оценкой в первом семестре принимается в форме проекта.

Для зачета необходимо выполнить следующее задание:

Реализовать программный код средней сложности на языке Python по текстовому описанию.

Проект сдается в письменном виде в соответствующем элементе системы Moodle. При сдаче проекта оценивается выполняемость программного кода, демонстрация понимания основных функций и структур языка Python, умение оценивать и применять программные методы в соответствии с поставленной задачей.

Результаты зачета с оценкой определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» ставится при своевременной и успешной сдаче проекта, своевременном выполнении домашних заданий и посещаемости не менее 80% занятий.

Оценка «хорошо» ставится при успешной сдаче проекта и выполнении домашних заданий, посещаемости не менее 60% занятий.

Оценка «удовлетворительно» ставится при сдаче проекта и выполнении домашних заданий при посещаемости менее 60% занятий.

Оценка «неудовлетворительно» ставится при несдаче проекта или невыполнении домашних заданий.

Итоговый проект проверяет компетенции ИПК-1.3, ИПК-3.1, ИПК-3.2, ИПК-4.1, ИПК-4.3.

## **11. Учебно-методическое обеспечение**

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=30939>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) Методические указания по организации самостоятельной работы студентов.

## **12. Перечень учебной литературы и ресурсов сети Интернет**

а) основная литература:

— Шолле Ф. Глубокое обучение на Python / Франсуа Шолле ; [пер. с англ. А. Киселев]. - Санкт-Петербург [и др.] : Питер, 2019. - 397 с.: ил. - ( Серия "Библиотека программиста" )

— Рашка С. Python и машинное обучение : машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow : пер. с англ. / Себастьян Рашка, Вахид Мирджалили. - 2-е изд., полностью пересм. и доп.. - Санкт-Петербург [и др.] : Диалектика, 2019. - 654 с.: ил.

б) дополнительная литература:

– Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Петер Флах ; [пер. с англ. А. А. Слинкина]. - Москва : ДМК Пресс, 2015. - 399 с.: ил., табл.

– Handbook of natural language processing / edited by Nitin Indurkha, Fred J. Damerau. - 2nd edition. - Boca Raton [a. o.] : Taylor & Francis Group, CRC Press, 2010. - 1 online resource (XXIII, 666 p.): ill. - ( Chapman & Hall/CRC machine learning & pattern recognition series ) . URL: <http://sun.tsu.ru/limit/2018/000623863/000623863.pdf>

– Лутц М. Изучаем Python/ М. Лутц – СПб.: Символ-Плюс, 2011. - 1280 с.

– Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020. – 313 p.

в) ресурсы сети Интернет:

– открытые онлайн-курсы

– документация языка программирования Python (<https://docs.python.org/3/>)

– документация библиотек языка программирования Python sklearn, Keras, Tensorflow и т.п.

## **13. Перечень информационных технологий**

а) лицензионное и свободно распространяемое программное обеспечение:

– интернет-браузер (Google Chrome, Yandex, и т. д.)

– публично доступный сервис для исполнения скриптов на языке Python Google Colab (<https://colab.research.google.com>)

– среда разработки для языка программирования Python Jupyter Notebook.

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

- б) информационные справочные системы:
- Электронный каталог Научной библиотеки ТГУ –  
<http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
  - Электронная библиотека (репозиторий) ТГУ –  
<http://vital.lib.tsu.ru/vital/access/manager/Index>

#### **14. Материально-техническое обеспечение**

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

#### **15. Информация о разработчиках**

Аишева Динара Армановна, ассистент каф. общей, компьютерной и когнитивной лингвистики,