

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

Филологический факультет

УТВЕРЖДЕНО:
Декан
И. В. Тубалова

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
2025

СОГЛАСОВАНО:
Руководитель ОП
З.И. Резанова

Председатель УМК
Ю.А. Тихомирова

Томск – 2025

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен решать профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий и актуальные концепции в области лингвистики.

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

ПК-3 Способен 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 Способен 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:

ИОПК-1.2 Решает профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий

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 Блоку 1 «Дисциплины (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

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: 10 ч.

-практические занятия: 32 ч.

в том числе практическая подготовка: 32 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 останова 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.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет с оценкой во втором семестре принимается в форме проекта.

Для зачета необходимо выполнить следующее задание:

Реализовать программный код средней сложности на языке Python по текстовому описанию следующего содержания:

«Примените один из рассмотренных в ходе курса методов машинного обучения для решения задачи классификации на текстовых данных, которые не рассматривались в ходе

курса (собранные самостоятельно в ходе курса Язык программирования Python или найденные в интернете).

Опишите задачу и объясните выбор метода решения задачи в комментариях к коду. Оцените получившийся классификатор с помощью метрик, изученных в ходе курса.

Прикрепите архив с данными или ссылку на них.»

Проект сдается в письменном виде в соответствующем элементе системы LMS iDo. 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 условиях: классификатор реализован без ошибок или с минимальными недочетами, не влияющими на результат, решает задачу с высокой точностью, код структурирован, содержит подробные комментарии с четким описанием задачи и обоснованием метода; метрики корректно применены и интерпретированы; своевременное выполнение домашних заданий и посещаемость не менее 80% занятий;

Оценка «хорошо» ставится при условиях: классификатор работает с незначительными недочетами, код читаем, комментарии объясняют задачу и выбор метода, но могут быть недостаточно подробными; метрики применены, но их интерпретация недостаточна; выполнение домашних заданий, посещаемость не менее 60% занятий.

Оценка «удовлетворительно» ставится при условиях: классификатор решает задачу с заметными недочетами, код минимально структурирован, комментарии недостаточны; метрики применены с ошибками или без интерпретации; выполнение домашних заданий, посещаемость менее 60% занятий.

Оценка «неудовлетворительно» ставится при условиях: проект не сдан, классификатор не работает или не решает задачу, код отсутствует, неструктурирован и не содержит комментариев; метрики не применены и не интерпретированы; домашние задания не выполнены.

Итоговый проект проверяет компетенции ИОПК-1.2, ИПК-1.3, ИПК-3.1, ИПК-3.2, ИПК-4.1, ИПК-4.3.

Продолжительность зачета с оценкой 1 час.

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «LMS iDo» - <https://lms.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 - <https://lms.tsu.ru/mod/resource/view.php?id=752567>

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;

Шамигов Федор Федорович, специалист по учебно-методической работе.