

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

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:
Директор
А. В. Замятин

Рабочая программа дисциплины

Введение в интеллектуальный анализ данных

по направлению подготовки

02.03.02 Фундаментальная информатика и информационные технологии

Направленность (профиль) подготовки:

Искусственный интеллект и разработка программных продуктов

Форма обучения
Очная

Квалификация
Бакалавр

Год приема
2025

СОГЛАСОВАНО:
Руководитель ОП
А.В. Замятин

Председатель УМК
С.П. Сущенко

Томск – 2025

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ПК-3. Способен осуществлять научно-исследовательские и опытно-конструкторские разработки как при исследовании самостоятельных тем, так и разработки по тематике организации.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИПК-3.1. Осуществляет проведение работ по обработке и анализу научно-технической информации и результатов исследований.

2. Задачи освоения дисциплины

– Освоить аппарат моделей и методов интеллектуального анализа данных, включая методы предварительной обработки данных регрессионные модели, нейросетевые модели, деревья решений, метод опорных векторов и др., а также ансамблевые подходы и метрики измерения различных расстояний и метрики оценки точности моделей.

– Научиться применять понятийный аппарат моделей и методов интеллектуального анализа данных для решения различных практических задач профессиональной деятельности.

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. Дисциплина входит в модуль «Искусственный интеллект».

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

Пятый семестр, зачет с оценкой

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам:

«Методы статистического анализа данных».

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 32 ч.

-лабораторные: 16 ч.

в том числе практическая подготовка: 16 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение. Основные понятия. Терминология. Области и примеры применения.

Тема 2. Этапы Data Science. Машинное обучение, общая постановка задачи. CRISP-DM. Регрессия, переобучение.

Тема 3. Топологии нейросетей и задачи для них. Нейросетевая классификация, Deep Learning. Сверточные нейронные сети.

Тема 4. Кластеризация (k-means). Метрики расстояний. Критерии точности (Карра, ROC, RMSE), ошибки I/II рода, гипотеза A/B.

Тема 5. Предварительная обработка данных. Оптимизация признакового пространства

Тема 6. Классификация (деревья решений). Классификация (статистическая, байесовский подход). SVM (метод опорных векторов). Регуляризация (L1, L2).

Тема 7. Ассоциативные алгоритмы (ассоциация, последовательная ассоциация).

Тема 8. Высокопроизводительная обработка данных (принципы и модели). Критерии эффективности.

Тема 9. Многоуровневое машинное обучение. Визуализация.

Тема 10. Обработка естественного языка. Программные среды и сервисы (Hadoop, MapReduce, Spark, Yarn, Cassandra).

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится на основе контроля посещаемости, подготовки и защиты рефератов, а также по результатам выполнения лабораторных работ 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 в пятом семестре проводится в смешанной форме (письменно-устной) по билетам. Экзаменационный билет состоит из трех частей. Продолжительность зачета с оценкой 1 час.

Критерии оценивания:

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

Оценка «отлично» выставляется, если даны правильные ответы на все вопросы без ошибок.

Оценка «хорошо» выставляется, если имеются незначительные неточности в ответах или незначительный дефицит в детализации ответа.

Оценка «удовлетворительно» выставляется, если имеются значительные неточности в ответах или значительный дефицит в детализации ответа.

Оценка «неудовлетворительно» выставляется, если отсутствует понимание предмета.

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 LMS iDo - <https://lms.tsu.ru/enrol/index.php?id=35022>

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

в) Методические указания по проведению лабораторных работ.

г) Методические указания по организации самостоятельной работы студентов.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Замятин А.В. Интеллектуальный анализ данных: учеб. пособие. – Томск: Издательский Дом государственного университета, 2020. – 196 с.

– Замятин А.В. Введение в интеллектуальный анализ данных : учебное пособие /А. В. Замятин. - Томск : Издательский Дом Томского государственного университета , 2016. - 118 с. – URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000529594.-1>

– Pocket Data Mining electronic resource : Big Data on Small Devices / by Mohamed Medhat Gaber, Frederic Stahl, João Bártolo Gomes. - Cham : : Springer International Publishing : : Imprint: Springer, , 2014. - 108 p. – URL: <http://dx.doi.org/10.1007/978-3-319-02711.-1>

- Principles of Data Mining electronic resource /by Max Bramer. - London : Springer London : Imprint: Springer, 2013. - 440 p. – URL: <http://dx.doi.org/10.1007/978-1-4471-4884-5>

б) дополнительная литература:

– Principles of Data Mining electronic resource /by Max Bramer.Bramer, Max. London : : Springer London :: Imprint: Springer, 2013, XIV, 440 p. [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1007/978-1-4471-4884-5>.

– Pocket Data Mining electronic resource : Big Data on Small Devices / /by Mohamed Medhat Gaber, Frederic Stahl, João Bártolo Gomes.Gaber, Mohamed Medhat. Cham : : Springer International Publishing : : Imprint: Springer, , 2014. IX, 108 p. [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1007/978-3-319-02711-1>.

– Миркин Б. Г. Введение в анализ данных : учебник и практикум для бакалавриата и магистратуры : [для студентов вузов, обучающихся по инженерно-техническим, естественно-научным и экономическим направлениям и специальностям] /Б. Г. Миркин ; "Высшая школа экономики" Национальный исследовательский университет. – Москва : Юрайт , 2015. – 173 с.

– Кулаичев А.П. Методы и средства комплексного анализа данных : учебное пособие. – Москва : Форум [и др.] , 2014. – 511 с.

в) ресурсы сети Интернет:

– открытые онлайн-курсы (Школа анализа данных (Яндекс) – <https://yandexdataschool.ru/>, Техносфера (образовательный проект Mail.Ru Group и факультета ВМК МГУ имени М. В. Ломоносова) – <https://sphere.mail.ru/pages/index/>.)

– Примеры источников данных:

– <https://www.kaggle.com/> (портал мировых соревнований команд по DM)

– <http://www.rdatamining.com/resources/data> (база данных географических наименований, маршрутов авиа сообщений, маркетинговая информация и т.п.);

– <http://www.kdnuggets.com/2014/08/interesting-social-media-datasets.html>+ (база данных социальных и медиа данных);

– http://monitoring-crm.ru/baza_dannyh_goszakupok/ (база данных государственных закупок);

– <https://archive.ics.uci.edu/ml/datasets/Iris> (база данных графических изображений);

– <https://data.gov.ru/> (Портал открытых данных РФ).

– Общероссийская Сеть КонсультантПлюс Справочная правовая система.
<http://www.consultant.ru>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– C, C++, C#, Python, R-Studio, Rapid Miner, MS Azure;

б) профессиональные базы данных:

– Data Mining for Service electronic. Berlin, Heidelberg, Imprint: Springer, Springer eBooks VIII, 291 p. 2014 (edited by Katsutoshi Yada) [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1007/978-3-642-45252-9>

– Data Mining for Geoinformatics electronic resource : Methods and Applications / edited by Guido Cervone, Jessica Lin, Nigel Waters. New York, NY : : Springer New York : : Imprint: Springer, , 2014, 166 p. [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1007/978-1-4614-7669-6>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий лабораторного 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.

Аудитории для проведения занятий лекционного типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

15. Информация о разработчиках

Замятин Александр Владимирович, д-р техн. наук, профессор, заведующий кафедрой теоретических основ информатики ТГУ, директор ИПМКН.

Карев Святослав Васильевич, ассистент кафедры теоретических основ информатики ИПМКН ТГУ