

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

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:
Директор

А.В. Замятин
« 19 » мая 20 22 г.

Рабочая программа дисциплины

Статистические методы машинного обучения

по направлению подготовки / специальности

10.05.01 Компьютерная безопасность

Направленность (профиль) подготовки / специализация:

Анализ безопасности компьютерных систем

Форма обучения

Очная

Квалификация

Специалист по защите информации

Год приема

2022

Код дисциплины в учебном плане: Б1.В.ДВ.01.01.02

СОГЛАСОВАНО:

Руководитель ОП

 В.Н. Тренькаев

Председатель УМК

 С.П. Сущенко

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

– ОПК-1 – Способен оценивать роль информации, информационных технологий и информационной безопасности в современном обществе, их значение для обеспечения объективных потребностей личности, общества и государства.

– ОПК-2 – Способен применять программные средства системного и прикладного назначений, в том числе отечественного производства, для решения задач профессиональной деятельности.

– ОПК-3 – Способен на основании совокупности математических методов разрабатывать, обосновывать и реализовывать процедуры решения задач профессиональной деятельности.

– ПК-2 – Способен разрабатывать требования к программно-аппаратным средствам защиты информации компьютерных систем и сетей.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.1 Учитывает современные тенденции развития информационных технологий в своей профессиональной деятельности.

ИОПК-2.3 Формулирует предложения по применению программных средств системного и прикладного назначений, в том числе отечественного производства, используемых для решения задач профессиональной деятельности.

ИОПК-3.2 Осуществляет применение основных понятий, фактов, концепций, принципов математики и информатики для решения задач профессиональной деятельности.

ИОПК-3.3 Выявляет научную сущность проблем, возникающих в ходе профессиональной деятельности, и применяет соответствующий математический аппарат для их формализации, анализа и выработки решения.

ИПК-2.2 Разрабатывает математические модели, реализуемые в средствах защиты информации.

ИПК-2.3 Проводит исследования с целью нахождения наиболее целесообразных практических решений по обеспечению защиты информации.

2. Задачи освоения дисциплины

– Познакомить с основными методами машинного обучения, применяемыми при анализе данных в экономике, медицине, социологии и других областях.

– Научить решать задачи статистического анализа данных с помощью моделей машинного обучения.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор. Дисциплина входит в «Профессиональные модули по выбору» в модуль "Введение в искусственный интеллект".

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

Шестой семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Общая алгебра», «Геометрия», «Математический анализ», «Теория вероятностей», «Математическая статистика».

6. Язык реализации

Русский

7. Объем дисциплины

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

-лекции: 16 ч.

-практические занятия: 32 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в статистический анализ данных.

Типы данных. Графические и табличные способы представления данных. Предварительная обработка данных. Оценки параметров и числовых характеристик. Проверка статистических гипотез.

Тема 2. Критерии сравнения групп.

Параметрические критерии. t-критерий Стьюдента. Критерий Фишера. Дисперсионный анализ. Непараметрические критерии. Критерии Манна-Уитни, Вилкоксона, Краскала-Уолиса, Фридмана.

Тема 3. Корреляционный анализ.

Парный коэффициент корреляции Пирсона. Z-преобразование Фишера. Ранговая корреляция. Коэффициент Спирмена, Кендалла, конкордации Кендалла. Корреляционный анализ категоризованных данных. Анализ таблиц сопряженности. Коэффициент Крамера.

Тема 4. Регрессионный анализ.

Постановка задачи регрессионного анализа. Метод наименьших квадратов оценки параметров простой регрессии. Условия Гаусса-Маркова. Теорема Гаусса-Маркова. Оценки дисперсий. Проверка качества модели регрессии, Коэффициент детерминации, его интерпретация, общая адекватность модели. Нелинейные модели и линеаризация. Фиктивные переменные. Случай смещенного шума. Случай коррелированных наблюдений Гетероскедастичность. Мультиколлинеарность.

Тема 5. Задачи классификации.

Основные понятия и задачи классификации. Бинарная классификация и логистическая регрессия. Метрики качества. ROC-анализ.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине осуществляется на основании проверки практических заданий, выполняемых студентами на компьютерах в течение семестра.

Студенты получают у преподавателя или собирают самостоятельно статистические данные для дальнейшего анализа и построения математических моделей. Проводят предварительную обработку данных, выбирают адекватный метод анализа в зависимости от целей исследования и типов данных, реализуют выбранные методы на компьютере, делают выводы и интерпретацию полученных результатов.

Результаты текущего контроля фиксируются в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзаменационная оценка складывается из текущего посещения (не менее 80% занятий), в срок выполненных практических заданий и результатов тестирования (при онлайн обучении) или письменного коллоквиума по темам:

1. Типы данных и способы их представления.
2. Параметрические критерии сравнения групп.

3. Непараметрические критерии сравнения групп.
4. Корреляционный анализ количественных данных.
5. Ранговая корреляция.
6. Корреляционный анализ категоризованных данных.
7. Парная регрессии. Модель. МНК-оценки параметров.
8. Числовые характеристики оценок параметров парной регрессии.
9. Теорема Гаусса-Маркова для случая парной регрессии.
10. Проверка качества уравнения парной регрессии.
11. Нелинейные модели и линеаризация.
12. Случай смещенного шума.
13. Случай коррелированных гомоскедастичных наблюдений.
14. Случай некоррелированных гетероскедастичных наблюдений.
15. Мультиколлинеарность.
16. Фиктивные переменные.
17. Постановка задачи классификации.
18. Логистическая регрессия.
19. Метрики качества бинарного классификатора.
20. ROC-анализ.

Посещение и сданные практические задания являются условием для допуска к теоретической части. Оценка за теоретическую часть ставится на основании теста или письменного коллоквиума.

Тест из 15 вопросов. Максимум 30 баллов.

0-15	Неудовлетворительно
16-20	Удовлетворительно
21-25	Хорошо
26-30	Отлично

Письменный коллоквиум. Два вопроса.

Ответ не дан или дан неверно, имеются грубые ошибки в формулировках и выводах	Неудовлетворительно
Ответ дан, но не в полном объеме, имеются существенные недочеты	Удовлетворительно
Ответ дан практически полностью, имеются некоторые незначительные ошибки	Хорошо
Ответ дан в полном объеме, допускаются очень незначительные погрешности	Отлично

При недостаточном посещении в течение семестра или невыполненных в срок работах студент может получить на экзамене дополнительные вопросы по пропущенным темам или дополнительное задание по практике.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle»

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине (Приложение 1).

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Джеймс Г., Уиттон Д., Хастис Е., Тибширани Р. Введение в статистическое обучение с примерами на языке R М.: ДМК Пресс, 2016. – 450 с.
- Кабанова Т. В. Применение пакета R для решения задач прикладной статистики : учебное пособие : [для студентов и аспирантов университетов]. Томск : Издательский дом Том. гос. ун-та, 2019. – 124 с.
- Марголис Н. Ю., Кабанова Т. В. Прикладная статистика: учебно-методическое пособие. Ч. 1. Том. гос. ун-т, 2007. – 46 с.
- Марголис Н. Ю., Кабанова Т. В. Прикладная статистика: учебно-методическое пособие. Ч. 2. Том. гос. ун-т, 2007. – 58 с.

б) дополнительная литература:

- М. Кендалл, А. Стьюарт Статистические выводы и связи. Наука.: Физматлит, 1973. – 432 с.
- С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин Прикладная статистика. Классификация и снижение размерности. Финансы и статистика, 1989. – 608 с.
- Айвазян С. А, Мхитарян В. С. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. . Т. 1. ЮНИТИ-ДАНА, 2001. – 270 с.
- Айвазян С. А. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. . Т. 2, ЮНИТИ-ДАНА, 2001. – 432 с.

в) ресурсы сети Интернет:

- <http://statsoft.ru/#tab-STATISTICA-link>
- <https://www.r-project.org/>
- <http://www-01.ibm.com/software/ru/analytics/spss/index.html>
- <http://itmu.vsuet.ru/Posobija/MathCAD/gl13/index.htm#anc1323>
- <http://www.exponenta.ru/>
- Общероссийская Сеть КонсультантПлюс Справочная правовая система.
<http://www.consultant.ru>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- MS Windows,
- MS Office,
- Statistica,
- R, R Studio.

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>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения практических занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы, оснащенные компьютерной техникой с установленным соответствующим ПО и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Кабанова Татьяна Валерьевна, канд. физ.-мат. наук, доцент, доцент кафедры теории вероятностей и математической статистики института прикладной математики и компьютерных наук НИ ТГУ.