

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ

Директор института прикладной
математики и компьютерных наук

А.В. Замятин
« 11 » ноября 2021 г.



Статистические методы машинного обучения

рабочая программа дисциплины

Закреплена за кафедрой Учебный план	<i>теории вероятностей и математической статистики 09.03.03 Прикладная информатика, профиль «Разработка программного обеспечения в цифровой экономике»</i>
Форма обучения	<i>очная</i>
Общая трудоёмкость	<i>4 з.е.</i>
Часов по учебному плану	<i>144</i>
в том числе:	
аудиторная контактная работа	<i>54,7</i>
самостоятельная работа	<i>57,6</i>
Вид(ы) контроля в семестрах экзамен/зачет/зачет с оценкой	<i>Семестр 6 – экзамен</i>

Программу составила:
канд. физ.-мат. наук, доцент
доцент кафедры теории вероятностей
и математической статистики



Т.В. Кабанова

Рецензент:
канд. физ.-мат. наук, доцент,
доцент кафедры теории вероятностей
и математической статистики



Е.Ю. Лисовская

Рабочая программа дисциплины «Статистические методы машинного обучения» разработана в соответствии с самостоятельно устанавливаемым образовательным стандартом высшего образования – бакалавриат – федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Томский государственный университет» по направлению подготовки 09.03.03 – Прикладная информатика (Утвержден Ученым советом НИ ТГУ, протокол от 27.10.2021 г. № 08).

Рабочая программа одобрена на заседании кафедры теории вероятностей и математической статистики

Протокол от 02 июня 2021 г. № 07

Заведующий кафедрой теории вероятностей
и математической статистики,
д-р физ.-мат. наук, профессор



С.П. Моисеева

Рабочая программа одобрена на заседании учебно-методической комиссии института прикладной математики и компьютерных наук (УМК ИПМКН)

Протокол от 17 июня 2021 г. № 05

Председатель УМК ИПМКН,
д-р техн. наук, профессор



С.П. Сущенко

Цель освоения дисциплины

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

Научить студентов решать задачи статистического анализа данных с помощью моделей машинного обучения, начиная от формулирования исходных задач соответствующей предметной области на языке прикладной статистики, выбора методов решения и критериев качества полученных решений и заканчивая формулировкой полученных выводов на языке предметной области.

1. Место дисциплины в структуре ОПОП

Дисциплина «Статистические методы машинного обучения» относится к вариативной части Блока 1 «Дисциплины», входит в модуль «Введение в искусственный интеллект».

Пререквизиты дисциплины: «Алгебра и геометрия», «Математический анализ», «Математическая логика и теория алгоритмов», «Теория вероятностей», «Случайные процессы», «Вычислительная математика», «Основы математического моделирования», «Математическая статистика», «Имитационное моделирование», «Методы оптимизации и исследование операций».

Постреквизиты дисциплины: технологическая (проектно-технологическая) практика, преддипломная практика, подготовка к процедуре защиты и защита выпускной квалификационной работы.

2. Компетенции и результаты обучения, формируемые в результате освоения дисциплины

Таблица 1.

Компетенция	Индикатор компетенции	Код и наименование результатов обучения (планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций)
ПК-3. Способен осуществлять научно-исследовательские и опытно-конструкторские разработки как при исследовании самостоятельных тем, так и разработки по тематике организации.	ИПК-3.1. Осуществляет проведение работ по обработке и анализу научно-технической информации и результатов исследований.	ОР-3.1.1 владеет: •навыками решения практических задач и анализа полученных результатов; •навыками формализации и построения алгоритма решения исследовательских задач.

3. Структура и содержание дисциплины

3.1. Структура и трудоемкость видов учебной работы по дисциплине

Общая трудоемкость дисциплины составляет 4 зачетные единицы, 144 часа.

Таблица 2.

Вид учебной работы	Трудоемкость в академических часах	
	6 семестр	всего
Общая трудоемкость	144	144
Контактная работа:	54,7	54,7
Лекции (Л):	16	16
Практики (ПЗ)	32	32
Лабораторные работы (ЛР)		

Семинары (СЗ)		
Групповые консультации	2	2
Индивидуальные консультации	2,4	2,4
Промежуточная аттестация	2,3	2,3
Самостоятельная работа обучающегося:	89,3	89,3
- изучение учебного материала, публикаций	47,7	47,7
- подготовка к лабораторным/практическим занятиям/коллоквиумам	9,9	9,9
- подготовка к рубежному контролю по теме/разделу	31,7	31,7
Вид промежуточной аттестации (зачет, зачет с оценкой, экзамен)	Экзамен	Экзамен

3.2. Содержание и трудоемкость разделов дисциплины

Таблица 3.

Код занятия	Наименование разделов и тем и их содержание	Вид учебной работы, занятий, контроля	Се м е с т р	Часы в электро нной форме	Всего (час.)	Литература	Код (ы) результата(ов) обучения
	Раздел 1. Введение в статистический анализ и первичная статистическая обработка		6		10	1,2,3,4,5,6,7,8	ОР-3.1.1
1.1.	Задачи и этапы статистического анализа	Лекции	6		2		
1.2.	Типы и структуры данных	Лекции	6		2		
1.3.	Предварительная обработка данных	Практика	6		2		
1.4.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		4		
	Раздел 2. Критерии сравнения групп		6		12	2, 4, 7, 8	ОР-3.1.1, ОР-3.1.2 ОР-3.1.3
2.1.	Параметрические критерии	Лекции, практика	6		4		
2.2.	Непараметрические критерии	Лекции, практика	6		4		
2.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		4		
	Раздел 3. Корреляционный анализ		6		12	1,2,3,5,6,7,8	ОР-3.1.1, ОР-3.1.2 ОР-3.1.3
3.1.	Количественная корреляция	Лекции, практика	6		4		
3.2.	Ранговая корреляция	Лекции, практика	6		2		
3.3.	Корреляционный анализ количественных данных	Лекции, практика	6		2		
3.4.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		4		
	Раздел 4. Регрессионный анализ				16	1,2,3,5,7,8	ОР-3.1.1, ОР-3.1.2 ОР-3.1.3
4.1.	Парная регрессия	Лекции, практика	6		6		
4.2.	Множественная регрессия	Лекции, практика	6		6		

4.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		4		
	Раздел 5. Задачи классификации		6		23,9	2,4	ОП-3.1.1, ОП-3.1.2 ОП-3.1.3
5.1.	Однофакторный дисперсионный анализ	Лекции, практика	6		8		
5.2.	Двухфакторный дисперсионный анализ	Лекции, практика	6		6		
5.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		9,9		
	Подготовка к промежуточной аттестации в форме экзамена	СРС	6		31,7	1, 2, 3, 4, 5, 6, 7, 8	
	Прохождение промежуточной аттестации в форме экзамена	Э	6		2,3		

4. Образовательные технологии, учебно-методическое и информационное обеспечение для освоения дисциплины

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.

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	Неудовлетворительно
---------------------------------------	---------------------

грубые ошибки в формулировках и выводах	
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.

4.1. Рекомендуемая литература и учебно-методическое обеспечение

№ п/п	Авторы / составители	Заглавие	Издательство	Год издания, количество страниц
Основная литература				
1.	Джеймс Г., Уиттон Д., Хастис Е., Тибширани Р.	Введение в статистическое обучение с примерами на языке R	М.: ДМК Пресс	2016 г., 450 с.
2.	Кабанова Т.В.	Применение пакета R для решения задач прикладной статистики: учебное пособие: [для студентов и аспирантов университетов]	Томск: Издательский Дом Томского государственного университета	2019 г., 124 с.
3.	Марголис Н.Ю., Кабанова Т.В.	Прикладная статистика: учебно-методическое пособие. Ч. 1	Том. гос. ун-т	2007 г., 46 с.
4.	Марголис Н.Ю., Кабанова Т.В.	Прикладная статистика: учебно-методическое пособие. Ч. 2	Том. гос. ун-т	2007 г., 58 с.
Дополнительная литература				
5.	Кендалл, М. Стьюарт А.	Статистические выводы и связи	Наука. Физматлит	1973 г., 432 с.
6.	Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.	Прикладная статистика. Классификация и снижение размерности	Финансы и статистика	1989 г., 608 с.
7.	Айвазян С.А, Мхитарян В.С.	Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 1	ЮНИТИ-ДАНА	2001 г., 270 с.
8.	Айвазян С.А.	Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 2	ЮНИТИ-ДАНА	2001 г., 432 с.

4.2. Базы данных и информационно-справочные системы, в том числе зарубежные

- 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/>

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>

4.3. Перечень лицензионного и программного обеспечения

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

4.4. Оборудование и технические средства обучения

Для реализации дисциплины необходимы лекционные аудитории и аудитории для проведения практических занятий. Специальные технические средства (проектор, компьютер и т.д.) требуются для демонстрации материала в рамках изучаемых разделов. Вся основная и дополнительная литература, необходимая для самостоятельной работы и подготовки к экзамену, имеется в научной библиотеке ТГУ.

5. Методические указания обучающимся по освоению дисциплины

Для самостоятельной работы студенты используют материалы, выданные преподавателем на лекции, самостоятельно изучают предложенную литературу, а также дополнительные источники (интернет-ресурсы по соответствующим темам и пр.).

6. Преподавательский состав, реализующий дисциплину

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

7. Язык преподавания – русский язык.

4. Образовательные технологии, учебно-методическое и информационное обеспечение для освоения дисциплины

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.

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:

21. Типы данных и способы их представления.
22. Параметрические критерии сравнения групп.
23. Непараметрические критерии сравнения групп.
24. Корреляционный анализ количественных данных.
25. Ранговая корреляция.
26. Корреляционный анализ категоризованных данных.
27. Парная регрессии. Модель. МНК-оценки параметров.
28. Числовые характеристики оценок параметров парной регрессии.
29. Теорема Гаусса-Маркова для случая парной регрессии.
30. Проверка качества уравнения парной регрессии.
31. Нелинейные модели и линеаризация.
32. Случай смещенного шума.
33. Случай коррелированных гомоскедастичных наблюдений.
34. Случай некоррелированных гетероскедастичных наблюдений.
35. Мультиколлинеарность.
36. Фиктивные переменные.
37. Постановка задачи классификации.
38. Логистическая регрессия.
39. Метрики качества бинарного классификатора.
40. 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.

4.1. Рекомендуемая литература и учебно-методическое обеспечение

№ п/п	Авторы / составители	Заглавие	Издательство	Год издания, количество страниц
Основная литература				
1.	Джеймс Г., Уиттон Д., Хастис Е., Тибширани Р.	Введение в статистическое обучение с примерами на языке R	М.: ДМК Пресс	2016 г., 450 с.
2.	Кабанова Т.В.	Применение пакета R для решения задач прикладной статистики: учебное пособие: [для студентов и аспирантов университетов]	Томск: Издательский Дом Томского государственного университета	2019 г., 124 с.
3.	Марголис Н.Ю., Кабанова Т.В.	Прикладная статистика: учебно-методическое пособие. Ч. 1	Том. гос. ун-т	2007 г., 46 с.
4.	Марголис Н.Ю., Кабанова Т.В.	Прикладная статистика: учебно-методическое пособие. Ч. 2	Том. гос. ун-т	2007 г., 58 с.
Дополнительная литература				
5.	Кендалл, М. Стьюарт А.	Статистические выводы и связи	Наука. Физматлит	1973 г., 432 с.
6.	Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.	Прикладная статистика. Классификация и снижение размерности	Финансы и статистика	1989 г., 608 с.
7.	Айвазян С.А, Мхитарян В.С.	Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 1	ЮНИТИ-ДАНА	2001 г., 270 с.
8.	Айвазян С.А.	Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 2	ЮНИТИ-ДАНА	2001 г., 432 с.

4.2. Базы данных и информационно-справочные системы, в том числе зарубежные

- 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/>

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>

4.3. Перечень лицензионного и программного обеспечения

– MS Windows,

– MS Office,

– Statistica,

– R, R Studio.

4.4. Оборудование и технические средства обучения

Для реализации дисциплины необходимы лекционные аудитории и аудитории для проведения практических занятий. Специальные технические средства (проектор, компьютер и т.д.) требуются для демонстрации материала в рамках изучаемых разделов. Вся основная и дополнительная литература, необходимая для самостоятельной работы и подготовки к экзамену, имеется в научной библиотеке ТГУ.

5. Методические указания обучающимся по освоению дисциплины

Для самостоятельной работы студенты используют материалы, выданные преподавателем на лекции, самостоятельно изучают предложенную литературу, а также дополнительные источники (интернет-ресурсы по соответствующим темам и пр.).

6. Преподавательский состав, реализующий дисциплину

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

7. Язык преподавания – русский язык.