

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

2021 г.



Статистические методы машинного обучения

рабочая программа дисциплины

Закреплена за кафедрой Учебный план	<i>теории вероятностей и математической статистики 01.03.02 Прикладная математика и информатика, профиль «Математические методы в 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>89.3</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 кафедры теории вероятностей
и математической статистики

Е.Ю. Лисовская

Рабочая программа дисциплины «Статистические методы машинного обучения» разработана в соответствии с образовательным стандартом высшего образования – бакалавриат, самостоятельно устанавливаемым федеральным государственным автономным образовательным учреждением высшего образования «Национальный исследовательский Томский государственный университет» по направлению подготовки 01.03.02 Прикладная математика и 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

И.о. заведующего 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 части, формируемой участниками образовательных отношений Блока «Факультативные дисциплины», входит в профессиональный модуль «Введение в искусственный интеллект».

Пререквизиты дисциплины: «Математический анализ», «Линейная алгебра и аналитическая геометрия», «Теория вероятностей», «Математическая статистика», «Методы оптимизации».

Постреквизиты дисциплины: НИР.

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)
ОПК-3. Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности	ИОПК-3.1. Демонстрирует навыки применения современного математического аппарата для построения адекватных математических моделей реальных процессов, объектов и систем в своей предметной области.	ОР-3.1.1. Знает основные методы статической обработки данные, умеет строить модели реальных процессов в различных прикладных областях. ОР-3.1.2. Владеет различными программными инструментами статистической обработки данных.
	ИОПК-3.2. Демонстрирует умение собирать и обрабатывать статистические, экспериментальные, теоретические и т.п. данные для построения математических моделей, расчетов и конкретных практических выводов.	ОР-3.2.1. Умеет собирать данные в соответствии с поставленной задачей конкретной предметной области и осуществлять их предобработку. ОР-3.2.2. Строит математические модели, осуществляет их анализ качества и интерпретирует полученные результаты.

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	
	5 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	29,6	29,6
- подготовка к лабораторным/практическим занятиям/коллоквиумам	28	28
- 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	С е м е с т р	Часы в электронной форме	Всего (час.)	Литература	Код (ы) результата(ов) обучения
	Раздел 1. Введение в статистический анализ.		6		20	1, 2, 3, 4, 5, 6, 7, 8	ОР-3.2.1
1.1.	Типы данных. Графические и табличные способы представления данных. Предварительная обработка данных. Оценки параметров и числовых характеристик. Проверка статистических гипотез.	Л, ПЗ	6		8		
1.2.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		12		
	Раздел 2. Критерии сравнения групп		6		22	2, 4, 7, 8	ОР-3.1.1, ОР-3.1.2, ОР-3.2.1, ОР-3.2.2
2.1.	Параметрические критерии. t-критерий Стьюдента. Критерий Фишера. Дисперсионный анализ. Непараметрические критерии. Критерии Манна-Уитни, Вилкоксона, Краскала-Уолиса, Фридмана.	Л, ПЗ	6		10		
2.2.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		12		
	Раздел 3. Корреляционный анализ		6		22	1, 2, 3, 5, 6, 7, 8	ОР-3.1.1, ОР-3.1.2, ОР-3.2.1, ОР-3.2.2
3.1.	Парный коэффициент корреляции Пирсона. Z-преобразование Фишера. Ранговая корреляция. Коэффициент Спирмена, Кендалла, конкордации Кендалла. Корреляционный анализ категоризованных данных. Анализ таблиц сопряженности. Коэффициент Крамера.	Л, ПЗ	6		10		
3.2.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		12		
	Раздел 4. Регрессионный анализ				24	1, 2, 3, 5, 7, 8	ОР-3.1.1, ОР-3.1.2, ОР-3.2.1, ОР-3.2.2
4.1.	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.	Л, ПЗ	6		12		
4.2.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		12		

	Раздел 6. Задачи классификации		6		17,6	1, 2, 6	ОП-3.1.1, ОП-3.1.2, ОП-3.2.1, ОП-3.2.2
6.1.	Основные понятия и задачи классификации. Бинарная классификация и логистическая регрессия. Метрики качества. ROC-анализ.	Л, ПЗ	6		8		
6.2.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		9,6		
	Консультации в период теоретического обучения	К	6		4,4		
	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

Изучение дисциплины осуществляется посредством изучения материалов на лекциях и практических занятиях, а также выполнения самостоятельных работ.

Образовательные технологии – классические лекции, практические занятия, «перевернутый класс», PBL.

Самостоятельная работа включает в себя изучение учебного материала, публикаций; подготовку к практическим занятиям, подготовку к рубежному контролю и др.

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.

Промежуточная аттестация: 6 семестр – экзамен.

Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих этапы формирования компетенций, и методические материалы, определяющие процедуры оценивания результатов обучения, приведены в Приложении 1 к рабочей программе «Фонд оценочных средств».

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., Хасты 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.	М. Кендалл, А. Стьюарт	Статистические выводы и связи	Наука. Физматлит	1973 г., 432 с.
6.	С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин	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	ЮНИТИ-ДАНА	2001 г., 270 с.

		специальностей вузов: В 2 т. . Т. 1		
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

1. <http://statsoft.ru/#tab-STATISTICA-link>
2. <https://www.r-project.org/>
3. <http://www-01.ibm.com/software/ru/analytics/spss/index.html>
4. <http://itmu.vsuet.ru/Posobija/MathCAD/gl13/index.htm#anc1323>
5. <http://www.exponenta.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.