

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 02.03.02 Фундаментальная информатика и информационные технологии Направленность (профиль) «Искусственный интеллект и разработка программных продуктов»</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>6 семестр – 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 образовательным стандартом высшего образования – бакалавриат, самостоятельно устанавливаемым федеральным государственным автономным образовательным учреждением высшего образования «Национальный исследовательский Томский государственный университет» по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии (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 1 «Дисциплины» входит в модуль «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», «Методы оптимизации и исследование операций», «Интеллектуальные системы», «Введение в интеллектуальный анализ данных».

Постреквизиты дисциплины: «Прикладные аспекты машинного обучения» «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)
ОПК-2. Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности	ИОПК-2.1. Обладает необходимыми знаниями основных концепций современных вычислительных систем	ОР-2.1.1 <b>знает</b> : <ul style="list-style-type: none"><li>• методы подбора и обработки информации относительно выбранной темы исследования;</li><li>• методы построения и анализа математических моделей процессов, относящихся к различным предметным областям;</li></ul>
	ИОПК-2.2. Использует методы высокопроизводительных вычислительных технологий, современного программного обеспечения, в том числе отечественного происхождения	ОР-2.2.1 <b>умеет</b> : <ul style="list-style-type: none"><li>• находить в специальной литературе необходимую информацию по соответствующей проблеме;</li></ul>
	ИОПК-2.3. Использует инструментальные средства высокопроизводительных вычислений в научной и практической деятельности	ОР-2.3.1 <b>умеет</b> : <ul style="list-style-type: none"><li>• выбрать метод исследования соответствующий поставленной цели с учетом априорных знаний об исследуемом объекте;</li></ul>

ПК-2. Способен проектировать базы данных, разрабатывать компоненты программных систем, обеспечивающих работу с базами данных, с помощью современных инструментальных средств и технологий	ИПК-2.2. Готов осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий	ОР-2.2.1 владеет: <ul style="list-style-type: none"> <li>• навыками решения практических задач и анализа полученных результатов;</li> <li>• навыками формализации и построения алгоритма решения исследовательских задач.</li> </ul>
---	---	--

### 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
<b>Общая трудоемкость</b>	<b>144</b>	<b>144</b>
<b>Контактная работа:</b>	<b>54.7</b>	<b>54.7</b>
Лекции (Л):	16	16
Практики (ПЗ)	32	32
Групповые консультации	2	2
Индивидуальные консультации	2.4	2.4
Промежуточная аттестация	2.3	2.3
<b>Самостоятельная работа обучающегося:</b>	<b>89.3</b>	<b>89.3</b>
- 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
<b>Вид промежуточной аттестации (зачет, зачет с оценкой, экзамен)</b>	<b>Экзамен</b>	<b>Экзамен</b>

### 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
	<b>Раздел 1. Введение в статистический анализ и первичная статистическая обработка</b>		<b>6</b>		<b>16</b>	<b>1, 2, 3, 4, 5, 6, 7, 8</b>	ОР-2.1.1
1.1.	Задачи и этапы статистического анализа	Л	6		2		
1.2.	Типы и структуры данных	Л	6		2		
1.3.	Предварительная обработка данных	ПЗ	6		2		
1.4.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		10		
	<b>Раздел 2. Критерии сравнения групп</b>		<b>6</b>		<b>18</b>	<b>2, 4, 7, 8</b>	ОР-2.1.1, ОР-2.2.1, ОР-2.3.1
2.1.	Параметрические критерии	Л, ПЗ	6		4		
2.2.	Непараметрические критерии	Л, ПЗ	6		4		
2.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		10		
	<b>Раздел 3. Корреляционный анализ</b>		<b>6</b>		<b>18</b>	<b>1, 2, 3, 5, 6, 7, 8</b>	ОР-2.1.1, ОР-2.2.1, ОР-2.3.1
3.1.	Количественная корреляция	Л, ПЗ	6		4		
3.2.	Ранговая корреляция	Л, ПЗ	6		2		
3.3.	Корреляционный анализ количественных данных	Л, ПЗ	6		2		
3.4.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		10		
	<b>Раздел 4. Регрессионный анализ</b>				<b>22</b>	<b>1, 2, 3, 5, 7, 8</b>	ОР-2.1.1, ОР-2.2.1, ОР-2.3.1
4.1.	Парная регрессия	Л, ПЗ	6		6		
4.2.	Множественная регрессия	Л, ПЗ	6		6		
4.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		10		
	<b>Раздел 5. Дисперсионный анализ</b>		<b>6</b>		<b>18</b>	<b>2, 4</b>	ОР-2.1.1, ОР-2.2.1, ОР-2.3.1
5.1.	Однофакторный дисперсионный анализ	Л, ПЗ	6		4		
5.2.	Двухфакторный дисперсионный анализ	Л, ПЗ	6		4		

5.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		10		
	<b>Раздел 6. Задачи классификации и кластеризации</b>		<b>6</b>		<b>13.6</b>	<b>1, 2, 6</b>	ОП-2.1.1, ОП-2.2.1, ОП-2.3.1
6.1.	Методы классификации	Л, ПЗ	6		4		
6.2.	Методы кластеризации	Л, ПЗ	6		2		
6.3.	Форма СРС (изучение учебного материала, публикаций; подготовка к лабораторным; подготовка к рубежному контролю по разделу)	СРС	6		7.6		
	Консультации в период теоретического обучения	Консультация	6		4.4		
	<b>Подготовка к промежуточной аттестации в форме экзамена</b>	СРС	<b>6</b>		<b>31,7</b>	<b>1, 2, 3, 4, 5, 6, 7, 8</b>	
	<b>Прохождение промежуточной аттестации в форме экзамена</b>	Э	<b>6</b>		<b>2.3</b>		

#### 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.

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., Уитгон 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.	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 специальностей вузов: В 2 т. . Т. 1	ЮНИТИ-ДАНА	2001 г., 270 с.
8.	Айвазян С. А.	Прикладная статистика. Основы эконометрики :	ЮНИТИ-ДАНА	2001 г., 432 с.

		Учебник для экономических специальностей вузов: В 2 т. . Т. 2		
--	--	--	--	--

#### **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://itnu.vsuet.ru/Posobija/MathCAD/gl13/index.htm#anc1323>
5. <http://www.exponenta.ru/>

#### **4.3. Перечень лицензионного и программного обеспечения**

MS Windows, MS Office, Mathcad, 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 кафедры теории вероятностей и математической статистики ИПМКН ТГУ.

#### **7. Язык преподавания – русский язык.**