

А. А. ХАЛАФЯН

МЕТОДЫ  
МАШИННОГО ОБУЧЕНИЯ  
В DATA MINING  
ПАКЕТА  
STATISTICA



Горячая линия-Телеком



А. А. ХАЛАФЯН

# МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В DATA MINING ПАКЕТА STATISTICA

*Рекомендовано Ученым советом*

*Федерального государственного бюджетного образовательного учреждения высшего образования «Кубанской государственной университет» (КубГУ) в качестве учебного пособия для студентов, обучающихся по направлениям подготовки:*

*01.03.02 – Прикладная математика и информатика (бакалавриат);*

*01.04.02 – Прикладная математика и информатика (магистратура);*

*09.03.03 – Прикладная информатика*

Москва  
Горячая линия – Телеком  
2022

УДК 004.9:519.25(075.8)

ББК 32.973

X17

Рецензент: директор СтатСофт Россия, канд. физ.-мат. наук *В. П. Боровиков*

**Халафян А. А.**

**X17** Методы машинного обучения в Data Mining пакета STATISTICA. Учебное пособие для вузов. – М.: Горячая линия – Телеком, 2022. – 260 с.: ил.

**ISBN 978-5-9912-0975-5.**

В настоящее время, благодаря совершенствованию технологий сбора и хранения данных в различных областях человеческой деятельности накоплены огромные массивы разнородных данных – количественных, качественных, текстовых, ограниченного и неограниченного объема. Поэтому в дополнении к методам многомерного анализа, как правило, основанных на парадигме среднего, появились современные технологии анализа данных, в частности Data Mining – добычи данных, или интеллектуального анализа данных. Методы машинного обучения Data Mining являются составной частью искусственного интеллекта (ИИ), проникающего практически во все сферы человеческой деятельности. Но ИИ – это программный продукт, разработанный человеком, и эффективность его работы зависит в том, числе и от того насколько правильно применены методы машинного обучения.

В издании освещены методы машинного обучения: деревья решений – общие деревья классификации и регрессии, CHAD-модели, интерактивные деревья, стохастический градиентный бустинг, случайные леса регрессии и классификации; процедуры обучения – методы опорных векторов, k-ближайших соседей, наивный байесовский классификатор; автоматизированные нейронные сети и программа DATA MINER. Книга написана на основе курсов, читаемых автором в Кубанском государственном университете. При описании методов использовались версии пакета STATISTICA 10, 13 (Tibco, USA).

Для студентов, изучающих математические и технические дисциплины, а также аспирантов, преподавателей вузов, специалистов в области Data Science, научных работников различных направлений, занимающихся анализом данных. Простая и доступная для широкого круга читателей форма изложения делает возможным использование пособия для самостоятельного изучения методов машинного обучения, реализованных в Data Mining пакета STATISTICA.

**ББК 32.973**

Адрес издательства в Интернет [WWW.TECHBOOK.RU](http://WWW.TECHBOOK.RU)

ISBN 978-5-9912-0975-5

© А. А. Халафян, 2022

© КубГУ, 2022

© Научно-техническое издательство  
«Горячая линия – Телеком», 2022

# Оглавление

ВВЕДЕНИЕ .....	3
<b>1. МЕТОДЫ ДЕРЕВЬЯ РЕШЕНИЙ .....</b>	<b>5</b>
1.1. Общие деревья классификации и регрессии .....	9
1.2. SNAID-модели .....	32
1.3. Интерактивные деревья .....	46
1.4. Стохастический градиентный бустинг .....	51
1.5. Случайные леса регрессии и классификации .....	67
<b>2. ОБОБЩЕННЫЕ МЕТОДЫ КЛАСТЕРНОГО АНАЛИЗА .....</b>	<b>77</b>
<b>3. ПРОЦЕДУРЫ ОБУЧЕНИЯ .....</b>	<b>102</b>
3.1. Метод опорных векторов .....	102
3.2. Метод k-ближайших соседей .....	121
3.3. Наивный байесовский классификатор .....	128
<b>4. АВТОМАТИЗИРОВАННЫЕ НЕЙРОННЫЕ СЕТИ .....</b>	<b>136</b>
4.1. Классификация .....	136
4.2. Кластерный анализ .....	174
4.3. Временные ряды (регрессия) .....	187
4.4. Временные ряды (классификация) .....	208
4.5. Регрессия .....	217
<b>5. ПРОГРАММА DATA MINER .....</b>	<b>234</b>
Литература .....	257