

Министерство науки и высшего образования РФ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

**СОГЛАСОВАНО**

**Заведующий кафедрой**

**Кафедра прикладной  
математики и компьютерной  
безопасности (ПМКБ\_ИКИТ)**

наименование кафедры

подпись, инициалы, фамилия

«\_\_\_» \_\_\_\_\_ 20\_\_ г.

институт, реализующий ОП ВО

**УТВЕРЖДАЮ**

**Заведующий кафедрой**

**Кафедра прикладной математики  
и компьютерной безопасности  
(ПМКБ\_ИКИТ)**

наименование кафедры

**А.А. Кытманов**

подпись, инициалы, фамилия

«\_\_\_» \_\_\_\_\_ 20\_\_ г.

институт, реализующий дисциплину

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
ПРОДВИНУТЫЕ МЕТОДЫ  
АНАЛИЗА ДАННЫХ (ADVANCED  
METHODS OF DATA ANALYSIS)**

Дисциплина Б1.В.03 Продвинутые методы анализа данных (Advanced Methods of Data Analysis)

Направление подготовки / 01.04.02 Прикладная математика и  
специальность информатика, программа 01.04.02.09 Data  
Science and Mathematical Modeling 2020г

Направленность  
(профиль)

Форма обучения

очная

Год набора

2020

Красноярск 2021

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования с учетом профессиональных стандартов по укрупненной группе

010000 «МАТЕМАТИКА И МЕХАНИКА»

---

Направление подготовки /специальность (профиль/специализация)

Направление 01.04.02 Прикладная математика и информатика,  
программа 01.04.02.09 Data Science and Mathematical Modeling 2020г.

---

Программу  
составили

---

## **1 Цели и задачи изучения дисциплины**

### **1.1 Цель преподавания дисциплины**

Целью изучения дисциплины является ознакомление студентов с продвинутыми методами анализа данных.

### **1.2 Задачи изучения дисциплины**

Задачами изучения дисциплины являются:

- Ознакомление с задачами обучения на размеченных данных, ансамблями алгоритмов, нейросетями, метрическими и логическими моделями, основами анализа временных рядов;

- Выработка умений постановки задач анализа данных, обоснованного выбора и применения изученных методов в прикладных исследованиях

- Получение практических навыков программирования алгоритмов машинного обучения средствами языка Python.

1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

<b>УК-2:Способен управлять проектом на всех этапах его жизненного цикла.</b>
<b>УК-2.1:Формулирует на основе поставленной проблемы проектную задачу и способ ее решения через реализацию проектного управления.</b>
<b>УК-2.2:Разрабатывает концепцию проекта в рамках обозначенной проблемы: формулирует цель, задачи, обосновывает актуальность, значимость, ожидаемые результаты и возможные сферы их применения.</b>
<b>УК-2.3:Планирует необходимые ресурсы, в том числе с учетом их заменимости.</b>
<b>УК-2.4:Разрабатывает план реализации проекта.</b>
<b>УК-2.5:Осуществляет мониторинг хода реализации проекта, корректирует отклонения, вносит дополнительные изменения в план реализации проекта, уточняет зоны ответственности участников проекта.</b>
<b>УК-1:Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий.</b>
<b>УК-1.2:Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению.</b>
<b>УК-1.3:Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников.</b>
<b>ПК-4:Способен разрабатывать и внедрять новые методы и технологии исследования данных.</b>
<b>ПК-4.1:Знает: состояние и перспективы развития информационных технологий, технологий данных в России и в мире; современные и перспективные методы сбора, хранения и передачи данных; источники данных, интенсивность генерации данных источниками; технические средства и среды сбора, хранения и обработки данных; современные и перспективные средства визуализации и</b>

**интерпретации данных; исследование операций; машинное обучение; математическое моделирование; методы сравнительного анализа.**

**ПК-4.2:Способен проводить аналитические и поисковые исследования по тематике информационных технологий, технологий данных.**

1.4 Место дисциплины (модуля) в структуре образовательной программы

Базовые модели и методы машинного обучения

Системы хранения, обработки и управления данными

Базовые модели и методы машинного обучения (The Basics of Machine Learning)

Системы хранения, обработки и управления данными (Data Storage and Management Systems)

Анализ больших данных

Машинное обучение и криптография

Прикладные задачи анализа данных

выполнение и защита выпускной квалификационной работы

Анализ больших данных (Big Data)

Машинное обучение и криптография (Applications of Machine Learning in Cryptography)

Прикладные задачи анализа данных (Applied Data Analysis)

Выполнение и защита выпускной квалификационной работы (Final certification)

1.5 Особенности реализации дисциплины

Язык реализации дисциплины Английский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

## 2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Семестр
		2
<b>Общая трудоемкость дисциплины</b>	<b>5 (180)</b>	<b>5 (180)</b>
<b>Контактная работа с преподавателем:</b>	<b>2 (72)</b>	<b>2 (72)</b>
занятия лекционного типа	1 (36)	1 (36)
занятия семинарского типа		
в том числе: семинары		
практические занятия	1 (36)	1 (36)
практикумы		
лабораторные работы		
другие виды контактной работы		
в том числе: групповые консультации		
индивидуальные консультации		
иная внеаудиторная контактная работа:		
групповые занятия		
индивидуальные занятия		
<b>Самостоятельная работа обучающихся:</b>	<b>2 (72)</b>	<b>2 (72)</b>
изучение теоретического курса (ТО)		
расчетно-графические задания, задачи (РГЗ)		
реферат, эссе (Р)		
курсовое проектирование (КП)	Нет	Нет
курсовая работа (КР)	Нет	Нет
<b>Промежуточная аттестация (Экзамен)</b>	<b>1 (36)</b>	<b>1 (36)</b>

### 3 Содержание дисциплины (модуля)

#### 3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа (акад. час)	Занятия семинарского типа		Самостоятельная работа, (акад. час)	Формируемые компетенции
			Семинары и/или Практические занятия (акад. час)	Лабораторные работы и/или Практикумы (акад. час)		
1	2	3	4	5	6	7
1	Distance-based models	4	4	0	10	
2	Logical models	6	6	0	8	
3	Ensemble algorithms	8	8	0	12	
4	Deep learning	6	6	0	12	
5	Unsupervised learning	4	4	0	12	
6	Time series analysis	8	8	0	18	
Всего		36	36	0	72	

#### 3.2 Занятия лекционного типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	- k nearest neighbors	2	0	0
2	1	- Parzen-window estimation in classification problems	2	0	0
3	2	- Tree-based methods	2	0	0
4	2	- CART modelling	2	0	0
5	2	- Tree-pruning algorithm	2	0	0

6	3	- Bagging	2	0	0
7	3	- Boosting	2	0	0
8	3	- Random forest	4	0	0
9	4	- Neural networks	2	0	0
10	4	- Convolutional Neural Networks with TensorFlow and Keras	2	0	0
11	4	- Neural network ensembles	2	0	0
12	5	- Association Rules	2	0	0
13	5	- Cluster Analysis	2	0	0
14	6	- Autocorrelation and Periodic Movements	2	0	0
15	6	- Stationarity and Nonstationarity	1	0	0
16	6	- Trends. Volatility	1	0	0
17	6	- Transforming Time Series	2	0	0
18	6	- ARMA Models for Stationary Time Series	2	0	0
Итого			26	0	0

### 3.3 Занятия семинарского типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	- k nearest neighbors	2	0	0
2	1	- Parzen-window estimation in classification problems	2	0	0
3	2	- Tree-based methods	2	0	0
4	2	- CART modelling	2	0	0
5	2	- Tree-pruning algorithm	2	0	0
6	3	- Bagging	2	0	0
7	3	- Boosting	2	0	0
8	3	- Random forest	4	0	0
9	4	- Neural networks	2	0	0

10	4	- Convolutional Neural Networks with TensorFlow and Keras	2	0	0
11	4	- Neural network ensembles	2	0	0
12	5	- Assosiation Rules	2	0	0
13	5	- Cluster Analysis	2	0	0
14	6	- Autocorrelation and Periodic Movements	2	0	0
15	6	- Stationarity and Nonstationarity	1	0	0
16	6	- Trends. Volatility	1	0	0
17	6	- Transforming Time Series	2	0	0
18	6	- ARMA Models for Stationary Time Series	2	0	0
Всего			26	0	0

### 3.4 Лабораторные занятия

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
Всего					

## 5 Фонд оценочных средств для проведения промежуточной аттестации

Оценочные средства находятся в приложении к рабочим программам дисциплин.

## 6 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

6.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Neter J., Wasserman W., Kutner M. H.	Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs	Boston: Irwin, 1990
Л1.2	Dunn P. F.	Measurement and Data Analysis for Engineering and Science: научное издание	Boca Raton: CRC Press, 2010
6.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год



Л2.1	Galushkin A. I.	Neural Networks Theory: with 176 figures	New York: Springer-Verlag, 2007
Л2.2	Зарова Е. В.	Applied Multivariate Statistical Analysis: Presentations for Lecturing and Working Examples with R=Прикладной многомерный статистический анализ: Презентации для лекций и примеры решений с использованием пакета R: Учебное пособие на английском языке	Москва: ООО "Научно-издательский центр ИНФРА-М", 2016

## **8 Методические указания для обучающихся по освоению дисциплины (модуля)**

Для успешного усвоения дисциплины и допуска к экзамену необходимо:

- Прослушать все лекции или, в случае пропуска занятий, изучить из материал самостоятельно по рекомендуемым источникам литературы;
- Выполнить все 9 индивидуальных заданий и защитить преподавателю в форме собеседования;
- Выполнить групповой проект и защитить его в форме докладов на семинаре.

Экзамент сдается в устной форме. Каждый экзаменационный билет состоит из 2-х теоретических вопросов из списка вопросов к экзамену.

## **9 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)**

### **9.1 Перечень необходимого программного обеспечения**

9.1.1	Для выполнения индивидуальных заданий по программированию алгоритмов анализа данных необходима программная среда вычислений Python версии не младше, чем 3.0, а также редактор кода Jupiter Notebook
-------	--

### **9.2 Перечень необходимых информационных справочных систем**

9.2.1	1. Система электронного обучения Сибирского федерального университета ( <a href="https://e.sfu-kras.ru">https://e.sfu-kras.ru</a> )
9.2.2	2. электронные информационно-справочные ресурсы научной библиотеки СФУ ( <a href="http://bik.sfu-kras.ru">http://bik.sfu-kras.ru</a> )
9.2.3	3. Интернет-репозиторий статистических данных Data-Planet ( <a href="https://www.data-planet.com/">https://www.data-planet.com/</a> )
9.2.4	4. Интернет-репозиторий статистических данных на платформе kaggle.com. Режим доступа: <a href="https://www.kaggle.com/datasets">https://www.kaggle.com/datasets</a>

## **10 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)**

Учебные лаборатории и классы, оснащенные современными компьютерами, объединенными в локальные вычислительные сети с выходом в Интернет, а также периферийным и проекционным оборудованием.