

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 » _____ 2021 г.



Введение в интеллектуальный анализ данных

рабочая программа дисциплины

Закреплена за кафедрой Учебный план	<i>Теоретических основ информатики 02.03.02 Фундаментальная информатика и информационные технологии, профиль «Искусственный интеллект и разработка программных продуктов»</i>
Форма обучения	<i>очная</i>
Общая трудоёмкость	<i>3 з.е.</i>
Часов по учебному плану	<i>108</i>
в том числе:	
аудиторная контактная работа	<i>50,65</i>
самостоятельная работа	<i>57,35</i>
Вид(ы) контроля в семестрах	
экзамен/зачет/зачет с оценкой	<i>5 семестр – зачет с оценкой</i>

Программу составил:

д-р техн. наук, профессор,
заведующий кафедрой теоретических основ информатики

А.В. Замятин

Рецензент:

д-р техн. наук, профессор,
заведующий кафедрой прикладной информатики



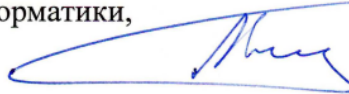
С.П. Сущенко

Рабочая программа дисциплины «Введение в интеллектуальный анализ данных» 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 теоретических основ информатики

Протокол от 04 июня 2021 г. № 05

Заведующий кафедрой теоретических основ информатики,
д-р техн. наук, профессор



А.В. Замятин

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

Целью освоения дисциплины «Введение в интеллектуальный анализ данных» является получение знаний в области моделей и методов интеллектуального анализа данных в задачах поиска информации, обработки и анализа данных, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

1. Место дисциплины в структуре ОПОП

Дисциплина «Введение в интеллектуальный анализ данных» относится к части, формируемой участниками образовательных отношений Блока 1 «Дисциплины», входит модуль «Искусственный интеллект».

Пререквизиты: «Интеллектуальные системы».

Постреквизиты: «Прикладные аспекты машинного обучения», «Нейронные сети», «Технологическая (проектно-технологическая) практика (стационарная)».

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

Таблица 1.

Компетенция	Индикатор универсальной компетенции	Код и наименование результатов обучения
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: Знать основные методы научно-практического поиска в задачах интеллектуального анализа данных и других областях с использованием информационных технологий. ОР-3.1.2: Знать существующие методы и подходы к интеллектуальному анализу данных различной природы. ОР-3.1.3: Уметь применять существующие методы интеллектуального анализа данных, обоснованно адаптируя и модифицируя их с учетом особенностей задачи предметной области.

3. Структура и содержание дисциплины

3.1. Структура и трудоемкость видов учебной работы по дисциплине

Общая трудоемкость дисциплины составляет 3 зачетные единицы, 108 часов.

Таблица 2.

Вид учебной работы	Трудоемкость в академических часах	
Общая трудоемкость	108	108
Контактная работа:	50,65	50,65
Лекции (Л):	32	32
Лабораторные работы (ЛР)	16	16
Групповые консультации	2,4	2,4
Промежуточная аттестация	0,25	0,25
Самостоятельная работа обучающегося:	57,35	57,35
- изучение учебного материала	18	18
- подготовка к лабораторным работам	22,1	22,1
- подготовка к рубежному контролю по дисциплине	17,25	17,25
Вид промежуточной аттестации	зачет с оценкой	зачет с оценкой

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	Коды результатов обучения
	Раздел 1. Основные проблемы построения систем		5		21,35	1	ОР-3.1.1, ОР-3.1.2.
1.1.	Актуальность, базовая терминология и тенденции развития. Основные задачи, этапы и классификация методов анализа данных.	Лекции	5		4		
1.2.	Предварительная обработка данных. Классификация.	Лекции	5		4		
1.3.	Регрессия. Ассоциация, последовательная ассоциация, аномалии и визуализация.	лабораторные	5		2		
1.4.	Высокопроизводительная обработка данных. Программные среды для интеллектуального анализа данных.	лабораторные	5		2		
1.5	Изучение учебного материала, подготовка к лабораторным работам, подготовка к рубежному контролю по дисциплине	СРС	5		9,35		
	Текущий контроль успеваемости		5				
	Раздел 2. Предварительная обработка данных. Классификация.		5		28	1, 2,3	ОР-3.1.1, ОР-3.1.2. ОР-3.1.3
2.1	Основные методы и предварительная обработка данных.	Лекции	5		4		
2.2	Оптимизация признакового пространства без трансформации пространства признаков.	Лекции	5		4		

2.3	Контролируемая непараметрическая нейросетевая классификация.	лабораторные	5		2		
2.4	Классификация по методу машины опорных векторов. Деревья решений.	лабораторные	5		2		
2.5	Изучение учебного материала, подготовка к лабораторным работам, подготовка к рубежному контролю по дисциплине	СРС	5		16		
	Текущий контроль успеваемости		5				
	Раздел 3. Регрессия. Ассоциация, последовательная ассоциация, аномалии и визуализация		5		28	1, 2,3	ОП-3.1.1, ОП-3.1.2. ОП-3.1.3
3.1	Понятие регрессии. Основные этапы регрессионного анализа.	Лекции	5		4		
3.2	Описание алгоритма ассоциации.	Лекции	5		4		
3.3	Алгоритмы семейства «Априори». Алгоритм GSP.	лабораторные	5		2		
3.4	Обнаружение аномалий и методы визуализации.	лабораторные	5		2		
3.5	Изучение учебного материала, подготовка к лабораторным работам, подготовка к рубежному контролю по дисциплине	СРС	5		16		
	Текущий контроль успеваемости		5				
	Раздел 4. Высокопроизводительная обработка данных. Программные среды для интеллектуального анализа данных		5		28	1, 2,3	ОП-3.1.1, ОП-3.1.2. ОП-3.1.3
4.1	Принципы организации высокопроизводительных вычислений. SMP-системы.	Лекции	5		4		
4.2	Модели параллельных вычислений MPMD, SPMD.	Лекции	5		4		
4.3	Платформа программирования и	лабораторные			4		

	выполнения распределённых вычислений Hadoop MapReduce, Mahout, Cassandra, Spark. Нереляционные базы данных HBase и язык NoSQL. Среды и языки программирования Python, R.1	торные	5				
4.5	Изучение учебного материала, подготовка к лабораторным работам, подготовка к рубежному контролю по дисциплине	СРС	5		16		
	Консультации в период теоретического обучения и промежуточной аттестации		5		2,4		
	Промежуточная аттестация	ЗаО	5		0,25		

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

Каждый студент реализует индивидуальный или групповой проект как последовательность лабораторных работ. Темы проектов имеют следующий шаблон:

1. Реализовать алгоритм анализа данных.
2. Предложить и реализовать технологии повышения производительности вычислений, выполняемых алгоритмом.

Самостоятельная работа студентов по предмету организуется в следующих формах:

- 1) самостоятельное изучение основного теоретического материала, ознакомление с дополнительной литературой, Интернет-ресурсами;
- 2) выполнение индивидуальных проектов, решение профессиональных задач из реальной предметной области.

Темы для изучения	Формы выполнения заданий	Количество часов
Актуальность, базовая терминология и тенденции развития. Основные задачи, этапы и классификация методов анализа данных	Самостоятельное изучение темы, предложенной преподавателем. Самостоятельное выполнение лабораторной работы №1	9,35
Предварительная обработка данных. Классификация.	Самостоятельное изучение темы, предложенной преподавателем. Самостоятельное выполнение лабораторной работы №2	16
Регрессия. Ассоциация, последовательная ассоциация, аномалии и визуализация	Самостоятельное изучение темы, предложенной преподавателем. Самостоятельное выполнение лабораторной работы №3	16
Высокопроизводительная обработка данных. Программные среды для интеллектуального анализа данных	Самостоятельное изучение темы, предложенной преподавателем. Самостоятельное выполнение лабораторной работы №4	16
Итого		57,35

В качестве учебно-методического обеспечения самостоятельной работы используется основная и дополнительная литература по предмету, Интернет-ресурсы, материал лекций, указания, выданные преподавателем при проведении лабораторных работ.

Типовые 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
Основная литература				
1	Замятин А.В.	А.В. Введение в интеллекту-	Издательский	2016, –194с.

		альный анализ данных	Дом государственного университета	
2	Mohamed Medhat Gaber, Frederic Stahl, João Bártolo Gomes.	Pocket Data Mining electronic resource : Big Data on Small Devices	Springer International Publishing : : Imprint: Springer	2014
3	Max Bramer	Principles of Data Mining electronic resource	Springer London : Imprint: Springer	2013
Дополнительная литература				
4	Миркин Б. Г.	Введение в анализ данных : учебник и практикум для бакалавриата и магистратуры : [для студентов вузов, обучающихся по инженерно-техническим, естественно-научным и экономическим направлениям и специальностям]	"Высшая школа экономики" Национальный исследовательский университет. – Москва: Юрайт	2015. – 173 с.
5	Кулаичев А.П.	Методы и средства комплексного анализа данных : учебное пособие.	Москва : Форум	2014. – 511 с.

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

1. Data Mining for Service electronic. Berlin, Heidelberg, Imprint: Springer, Springer eBooks VIII, 291 p. 2014 (edited by Katsutoshi Yada) [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1007/978-3-642-45252-9>
2. Data Mining for Geoinformatics electronic resource : Methods and Applications / /edited by Guido Cervone, Jessica Lin, Nigel Waters. New York, NY : : Springer New York : : Imprint: Springer, , 2014, 166 p. [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1007/978-1-4614-7669-6>

4.3. Перечень лицензионного и программного обеспечения

Средства и среды программирования C, C++, C#, Python, R-Studio, Rapid Miner, MS Azure.

4.4. Оборудование и технические средства обучения

Для материально-технического обеспечения дисциплины требуется наличие компьютерной техники с установленным соответствующим программным обеспечением и другого оборудования, поддерживающего проведение презентаций, построение проектной документации, выход в сеть Интернет.

5. Методические указания обучающимся по освоению дисциплины

Для успешного освоения дисциплины студенты должны посещать занятия, прорабатывать указанные материалы для самостоятельной работы студентов, выполнять лабораторные работы.

6. Преподавательский состав, реализующий дисциплину

Замятин Александр Владимирович, д-р техн. наук, профессор, заведующий кафедрой теоретических основ информатики ТГУ.

7. Язык преподавания – русский