

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
«Новосибирский национальный исследовательский государственный университет»  
(Новосибирский государственный университет, НГУ)  
**Структурное подразделение Новосибирского государственного университета –  
Специализированный учебно-научный центр Университета (СУНЦ НГУ)**  
Министерство науки и высшего образования Российской Федерации

СОГЛАСОВАНО Заместитель директора по УР  (Петровская О.В.) 23 ноября 2023 г.	УТВЕРЖДЕНО На заседании ученого совета СУНЦ НГУ Протокол № 48 от 23 ноября 2023 г.	УТВЕРЖДАЮ Директор СУНЦ НГУ  (Некрасова Л.А.) 23 ноября 2023 г.
---	--	--

**РАБОЧАЯ ПРОГРАММА**

**курса внеурочной деятельности «Введение в анализ данных»**

Заведующий кафедрой дискретной математики и информатики

Гончаров Сергей Савостьянович, д.ф.-м.н., академик РАН



Новосибирск 2023

## **ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

Спецкурс «Введение в анализ данных» направлен на расширение знаний учеников 9 - 11-х классов по информатике, физике и математике, способствующий расширению кругозора на парадигмы и возможности методов анализа данных и дальнейшей интерпретации полученных результатов. Программа спецкурса составлена с учетом специфики преподавания математики, физики и информатики в СУНЦ НГУ и предусматривает знакомство учащихся с методами анализа данных и областью их применения. Спецкурс строится с опорой на умения и навыки решения задач векторной алгебры, комплексных чисел, программирования алгоритмов.

Цель курса – развитие представления школьников о методах решения на ЭВМ задач анализа и интерпретации данных.

## **ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ**

В результате освоения программы спецкурса обучающийся должен:

- Изучить основные методы решения на ЭВМ задач анализа и интерпретации данных, получаемых с помощью различного рода информационно-измерительных систем.
- Изучить базовые алгоритмы анализа и интерпретации данных.
- Приобрести практические навыки работы с современными пакетами прикладных программ для решения задач анализа и интерпретации данных.

## **СОДЕРЖАНИЕ СПЕЦКУРСА**

Введение (2 часа) Предмет курса и его задачи. Краткие сведения о развитии теории анализа и интерпретации данных.

Тема 1. Основные понятия дисциплины (20 часов) Введение в анализ данных. Проблема обработки данных. Матрица данных. Гипотезы компактности и скрытых факторов. Структура матрицы данных и задачи обработки. Матрица объект-объект и признак-признак. Расстояние и близость. Измерение признаков. Отношения и их представление. Основные проблемы измерений. Основные типы шкал. Проблема адекватности. Основные задачи анализа и интерпретации данных.

Тема 2. Классификация данных с использованием детерминированных моделей (20 часов) Пространство весов. Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило. Обобщенные градиентные методы. Процедуры обучения на основе минимальной среднеквадратичной ошибки.

Тема 3. Классификация данных на основе статистических моделей (20 часов) Функция потерь. Принятие решение по максимуму правдоподобия. Множественный дискриминантный анализ. Пошаговый анализ. Ошибки классификации. Оценки максимального правдоподобия, байесовские оценки.

Тема 4. Кластер-анализ (20 часа) Основные типы задач кластер-анализа. Меры подобия и функции расстояния. Выбор критерия кластеризации. Кластерные методы, основанные на евклидовой метрике. Иерархическая кластеризация. Использование методов теории графов в задачах кластеризации.

Тема 5. Методы снижения размерностей данных (20 часов) Анализ матриц исходных данных. Метод главных компонент. Корреляционная матрица и ее основные свойства. Собственные векторы и собственные числа корреляционной матрицы. Приведение корреляционной матрицы к диагональной форме. Геометрическая интерпретация главных компонент на плоскости. Модели факторного анализа. Использование кластеризации признаков для снижения размерности. Многомерное шкалирование.

Тема 6. Системы DATA MINING в задачах анализа и интерпретации данных (20 часов) Понятие об интеллектуальных системах анализа и интерпретации данных. DATA MINING - системы извлечения новых знаний из данных. Типы систем DATA MINING - предметно-ориентированные аналитические системы, статистические пакеты, нейронные сети, деревья решений, обнаружение логических закономерностей, генетические алгоритмы, системы визуализации многомерных данных

Тема 7. Современные пакеты прикладных программ для решения задач обработки экспериментальных данных (20 часов) Табличные процессоры и базы данных в задачах обработки данных. Виды статистических пакетов. Требования к статистическим пакетам общего назначения.

Заключение (2 часа) Перспективные направления развития и использования методов анализа и интерпретации данных

## ТЕМАТИЧЕСКОЕ ПЛАНИРОВАНИЕ

№ п/п	Наименование разделов и тем программы	Кол-во часов	Воспитательный компонент
1.	Введение в анализ данных	2	<p>Готовность обучающихся к саморазвитию, самостоятельности и личному самоопределению.</p> <p>Сформированность ценностей самостоятельности и инициативы.</p> <p>Сформированность мировоззрения, соответствующего современному уровню развития науки, достижениям научно-технического прогресса и общественной практики, за счёт понимания роли информационных ресурсов, информационных процессов.</p>
2.	Классификация данных с использованием детерминированных моделей	6	
3.	Классификация данных на основе статистических моделей	6	
4.	Кластер-анализ	6	
5.	Методы снижения размерностей данных	6	
6.	Системы DATA MINING в задачах анализа и интерпретации данных	6	
7.	Современные пакеты прикладных программ для решения задач обработки экспериментальных данных	6	
8.	Заключение	2	
Всего		40	