



Федеральное агентство морского и речного транспорта
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Государственный университет морского и речного флота
имени адмирала С.О. Макарова»**
Воронежский филиал ФГБОУ ВО «ГУМРФ имени адмирала С.О. Макарова»

Кафедра математики, информационных систем и технологий

АННОТАЦИЯ

дисциплины *«Анализ больших данных»*

Направление подготовки 09.03.02 Информационные системы и технологии

Направленность (профиль) Информационные системы на транспорте

Уровень высшего образования бакалавриат

Промежуточная аттестация экзамен

1. Место дисциплины в структуре образовательной программы

Дисциплина «Анализ больших данных» относится к обязательной части Блока 1 и изучается на 3 курсе в 6 семестре п очной форме обучения и на 4 курсе в 7 и 8 семестрах по заочной форме обучения.

Изучение дисциплины основано на принципах дальнейшего развития математических дисциплин обязательной части программы, в том числе дисциплин «Математический анализ», «Теория информационных процессов и систем», «Моделирование процессов и систем», «Дифференциальные уравнения», «Автоматизация гидротехнических сооружений и водные пути»

В качестве «входных» знаний и умений требуется владение основными понятиями теории вероятностей, математической статистики, конечномерного линейного анализа, прежде всего операциями с матрицами и квадратичными формами.

Освоение данной дисциплины как предшествующей необходимо для последующего овладения дисциплинами «Основы Data Mining», «Методы искусственного интеллекта», «Технологии интеллектуального анализа данных», «Стандартизация и унификация информационных технологий» и для решения задач дипломного проектирования.

2. Планируемые результаты обучения по дисциплине

Код и наименование компетенции	Код индикатора достижения компетенции	Планируемые результаты обучения по дисциплине
ПК-3: Способность проводить научные исследования при разработке, внедрении и сопровождении информационных систем и технологий с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	ПК-3.1: анализ исходных данных, оценку качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	Знать: проводить анализ исходных данных, оценку качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Уметь: анализировать информационные системы и технологии с целью выявления показателей, требующих улучшения, с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Владеть: навыками оптимизации информационной системы в области анализа больших данных
	ПК-3.2 Оптимизация информационных систем и технологий для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	Знать: Инструменты и методы оценки качества и эффективности ИС Уметь: оптимизировать информационные системы и технологии для достижения новых целевых показателей с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Владеть: навыками осуществления оптимизации ИС для достижения новых целевых показателей в области анализа больших данных
ОПК-8: Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем.	ОПК-8.1. Математическое моделирование сложных систем, анализ данных	Знать: требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных Уметь: использовать современные инструментальные и вычислительные средства, осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов Владеть: способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям

3. Объем дисциплины по видам учебных занятий

Объем дисциплины составляет 6 зачетных единиц; всего 216 часов, из которых по очной форме 102 часа составляет контактная работа обучающегося с преподавателем (51 час – занятия лекционного типа, 51 час – лабораторные занятия), по заочной форме 32 часа составляет контактная работа обучающегося с преподавателем (16 часов– занятия лекционного типа, 16 часов –лабораторные занятия).

4. Основное содержание дисциплины

Информация и данные, системы и модели. Три уровня анализа информации. Информация и данные. Информатика и информационные системы. Интеллектуальный анализ данных (ИАД). Data Mining. Аналитические модели. Имитационные модели. Большие системы. Модели вместо законов. Многозначность понятий. Человек и машина. Теорема Гёделя. Генератор случая. Живое и неживое. Социо-кибер-физические системы.

Большие данные и инфраструктура их обработки. Процесс аналитики больших данных. Data Mining и элементы машинного обучения. Концепция MapReduce. Hadoop и Spark в экосистеме больших данных. Базы данных NoSQL. Интеллектуальный анализ текста как источника больших данных. Визуализация больших данных.

Классификация многомерных измерений и кластерный анализ. Дискриминантные информанты и классификация. Оценка вероятностей ошибочной классификации. Классификация на линейных дискриминантных формах. Кластеризация. Выбор метрики. Метод k средних и ЕМ-алгоритм. Иерархическая кластеризация на основе дендрограммы. Оценка качества разделения.

Рекуррентные и генетические алгоритмы, алгоритмы прямого поиска. m - мерный линейный фильтр Калмана. Варианты определения исходных данных. Решетчатые фильтры и МК-фильтры. Слияние разнородной информации.

Краткие сведения о генетических алгоритмах. Естественный отбор в природе. Особенности генетических алгоритмов. Структура генетического алгоритма. Графический интерфейс генетического алгоритма. Графические возможности интерфейса. Использование генетического алгоритма в режиме CLI. Использование алгоритма с опциями по умолчанию. Описание алгоритма поиска по образцу. Графический интерфейс алгоритма поиска по образцу. Использование алгоритма поиска по образцу в CLI.

Составитель: д.ф.-м.н., профессор Кузьменко Р. В.

Зав. кафедрой: к.ф.-м.н., доцент Черняева С. Н.