

Федеральное государственное бюджетное образовательное учреждение высшего образования «Тамбовский государственный университет имени Г.Р. Державина»  
Институт новых технологий и искусственного интеллекта  
Кафедра функционального анализа

УТВЕРЖДАЮ:  
И.о. директора института



Н. Л. Королева  
«16» сентября 2024 г.

## **РАБОЧАЯ ПРОГРАММА**

по дисциплине Б1.В.ДВ.04.2 Методы и системы обработки больших данных

Направление подготовки/специальность: 01.03.01 - Математика

Профиль/направленность/специализация: Искусственный интеллект и моделирование

Уровень высшего образования: бакалавриат

Квалификация: Бакалавр

год набора: 2024

**Авторы программы:**

Рыбаков Михаил Анатольевич

Кандидат физико-математических наук, Переславцева Оксана Николаевна

Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 01.03.01 - Математика (уровень бакалавриата) (приказ Министерства образования и науки РФ от «10» января 2018 г. № 8).

Рабочая программа принята на заседании Кафедры функционального анализа «13» сентября 2024 г. Протокол № 2

Рассмотрена и одобрена на заседании Ученого совета Института новых технологий и искусственного интеллекта, Протокол от «16» сентября 2024 г. № 1.

## СОДЕРЖАНИЕ

1. Цели и задачи дисциплины.....	4
2. Место дисциплины в структуре ОП Бакалавриата.....	4
3. Объем и содержание дисциплины.....	4
4. Контроль знаний обучающихся и типовые оценочные средства.....	6
5. Методические указания для обучающихся по освоению дисциплины (модуля).....	9
6. Учебно-методическое и информационное обеспечение дисциплины.....	11
7. Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы.....	12

## 1. Цели и задачи дисциплины

1.1 Цель дисциплины – формирование компетенций:

1.2 Типы задач профессиональной деятельности, к которым готовятся обучающиеся в рамках освоения дисциплины:

- научно-исследовательский

1.3 Дисциплина ориентирована на подготовку обучающихся к профессиональной деятельности в сфере: 40 Сквозные виды профессиональной деятельности в промышленности (в сфере научно-исследовательских и опытноконструкторских разработок)

1.4 В результате освоения дисциплины у обучающихся должны быть сформированы:

Обобщенные трудовые функции / трудовые функции / трудовые или профессиональные действия (при наличии профстандарта)	Код и наименование компетенции ФГОС ВО, необходимой для формирования трудового или профессионального действия	Индикаторы достижения компетенций
---	---	-----------------------------------

1.5 Согласование междисциплинарных связей дисциплин, обеспечивающих освоение компетенций:

## 2. Место дисциплины в структуре ОП бакалавриата:

Дисциплина «Методы и системы обработки больших данных» относится к части, формируемой участниками образовательных отношений, учебного плана ОП по направлению подготовки 01.03.01 - Математика.

Дисциплина «Методы и системы обработки больших данных» изучается в 4 семестре.

## 3. Объем и содержание дисциплины

3.1. Объем дисциплины: 2 з.е.

Очная: 2 з.е.

Вид учебной работы	Очная (всего часов)
<b>Общая трудоёмкость дисциплины</b>	<b>72</b>
Контактная работа	28
Лабораторные (Лаб. раб.)	28
Самостоятельная работа (СР)	44
Зачет	-

3.2. Содержание курса:

№ темы	Название раздела/темы	Вид учебной работы, час.		Формы текущего контроля
		Лаб	СР	
		раб.		
		О	О	
4 семестр				

1	Дynamo-подобные хранилища, Cassandra	10	16	Лабораторная работа; Опрос
2	Аналитические методы обработки больших данных.	10	16	Лабораторная работа; Опрос
3	Методы обработки больших данных на основе искусственных нейронных сетей	8	12	Лабораторная работа; Тестирование

### **Тема 1. Dynamo-подобные хранилища, Cassandra**

#### **Лекция.**

Модель данных Dynamo/Cassandra. Хеш-партиционирование данных, consistent hashing, eventual consistency, read quorum, write quorum. Антиэнтропийные техники. Hinted Handoff. Merkle Trees и их использование при синхронизации реплик.

#### **Лабораторные работы.**

#### **Лабораторное занятие.**

1. Установка и конфигурация Cassandra.

#### **Задания для самостоятельной работы.**

Задачи:

1. Подготовка к защите лабораторной работы.

### **Тема 2. Аналитические методы обработки больших данных.**

#### **Лекция.**

Аналитические методы обработки больших данных. Платформы прогнозной аналитики. RapidMiner. Этапы глубинного анализа данных. Результирующая визуализация. Проверка. Оптимизация. STATISTICA Data Miner.

Предварительная обработка данных. Фильтрация данных. Чистка данных

#### **Лабораторные работы.**

#### **Лабораторное занятие.**

1. Работа в GUI RapidMine.

2. Работа в STATISTICA Data Miner.

3. Генерация прогнозных моделей STATISTICA Data Miner в различных форматах.

#### **Задания для самостоятельной работы.**

Задачи:

1. Подготовка к защите лабораторной работы.

### **Тема 3. Методы обработки больших данных на основе искусственных нейронных сетей**

#### **Лекция.**

Понятия о нейронных сетях (НС). Общие черты НС. Нейроны. Классификация по типу нейронов: классификация нейронов по виду функции стимуляции, классификация нейронов по виду функции активации, классификация нейронов по типу сигнала, классификация нейронов по вероятностной определенности. Классификация нейросетевых архитектур: по типу связей в сети - сеть прямого распространения, рекуррентная сеть, соревновательная сеть; по наличию скрытых элементов; по динамическим характеристикам. Классификация методов обучения: по используемым парадигмам - «с учителем» (контролируемое обучение), «без учителя» (самообучение) и смешанная; по используемым правилам - коррекция по ошибке; машина Больцмана; правило Хебба; обучение методом соревнования; обучение методом селекции. Многослойный персептрон (MLP). Методы нахождения глобального минимума функции ошибок. Радиальная базисная функция (RBF). Преимущества сети RBF перед сетями MLP. Принципы обучения RBF-сети. Вероятностная нейронная сеть (PNN). Преимущества и недостатки PNN-сетей. Сеть Кохонена. Основной итерационный алгоритм Кохонена. Область применения сети Кохонена. Необходимость построения гибридной сети. Эффективность гибридизации. Принцип работы гибридной сети. Отличие ГС от обычной нейросети. Принцип обучения ГС. База знаний гибридной сети. Алгоритм вывода для ГС. Адаптивная нечеткая нейросистема (ANFIS). Архитектура нейронной сети ANFIS. Примеры использования системы ANFIS: построение регрессионно-авторегрессионной модели сильно зашумленного ряда; прогноз сильно зашумленного ряда.

### Лабораторные работы.

#### Лабораторное занятие.

1. Разработка структуры НС.
2. Обучение НС обработки больших данных.

#### Задания для самостоятельной работы.

Задачи:

1. Подготовка к защите лабораторной работы

## 4. Контроль знаний обучающихся и типовые оценочные средства

### 4.1. Распределение баллов:

4 семестр

- посещаемость – 10 баллов
- текущий контроль – 70 баллов
- контрольные срезы – 2 среза по 10 баллов каждый
- премиальные баллы – 20 баллов

#### Распределение баллов по заданиям:

№ те мы	Название темы / вид учебной работы	Формы текущего контроля / срезы	Мах. кол-во баллов	Методика проведения занятия и оценки
1.	Динамо-подобные хранилища, Cassandra	Лабораторная работа	25	В случае успешного выполнения всех заданий лабораторной работы студент получает 10 баллов
		Опрос	5	В случае успешного ответа на все вопросы студент получает 5 баллов
2.	Аналитические методы обработки больших данных.	Лабораторная работа	20	В случае успешного выполнения всех заданий лабораторной работы студент получает 10 баллов
		Опрос(контрольный срез)	10	В случае успешного ответа на все вопросы студент получает 10 баллов

3.	Методы обработки больших данных на основе искусственных нейронных сетей	Лабораторная работа	20	В случае успешного выполнения всех заданий лабораторной работы студент получает 10 баллов
		Тестирование(контрольный срез)	10	В случае правильных ответов на 51% заданий тестирования студент получает 10 баллов
4.	Посещаемость		10	10 баллов за выполнение в срок всех заданий текущего контроля и отсутствие пропусков занятий без уважительных причин
5.	Премиальные баллы		20	20 баллов за участие в студенческих научных конференциях и / или олимпиадах
6.	Индивидуальные задания, с помощью которых можно набрать дополнительные баллы		50	Студент может предоставить все задания текущего контроля
7.	Итого за семестр		100	

Итоговая оценка по зачету выставляется в 100-балльной шкале и в традиционной четырехбалльной шкале. Перевод 100-балльной рейтинговой оценки по дисциплине в традиционную четырехбалльную осуществляется следующим образом:

100-балльная система	Традиционная система
50 - 100 баллов	Зачтено
0 - 49 баллов	Не зачтено

#### 4.2 Типовые оценочные средства текущего контроля

### Лабораторная работа

#### Тема 1. Dynamo-подобные хранилища, Cassandra

#### Задания к лабораторной работе по теме №4 "Dynamo-подобные хранилища, Cassandra".

##### 1. Установка и конфигурация Cassandra.

#### Тема 2. Аналитические методы обработки больших данных.

#### Задания к лабораторной работе по теме №5 "Аналитические методы обработки больших данных".

##### 1. Работа в GUI RapidMine.

##### 2. Работа в STATISTICA Data Miner.

##### 3. Генерация прогнозных моделей STATISTICA Data Miner в различных форматах.

#### Тема 3. Методы обработки больших данных на основе искусственных нейронных сетей

#### Задания к лабораторной работе по теме №6 "Методы обработки больших данных на основе искусственных нейронных сетей".

##### 1. Разработка структуры НС.

##### 2. Обучение НС обработки больших данных.

### Опрос

#### Тема 1. Dynamo-подобные хранилища, Cassandra

Вопросы для проведения опроса по теории

- 1 Модель данных Dynamo/Cassandra.
- 2 Хеш-партиционирование данных, consistent hashing, eventual consistency, read quorum, write quorum.
- 3 Антиэнтропийные техники.
- 4 Hinted Handoff.
- 5 Merkle Trees и их использование при синхронизации реплик.

## Тема 2. Аналитические методы обработки больших данных.

Вопросы для проведения теоретического опроса в письменной форме

- 1 Аналитические методы обработки больших данных.
- 2 Платформы прогнозной аналитики.
- 3 RapidMiner.
- 4 Этапы глубинного анализа данных.
- 5 Результирующая визуализация.
- 6 Проверка.
- 7 Оптимизация.
- 8 STATISTICA Data Miner.
- 9 Предварительная обработка данных.
- 10 Фильтрация данных.
- 11 Чистка данных

## Тестирование

### Тема 3. Методы обработки больших данных на основе искусственных нейронных сетей

#### Тестовые задания

- 1 Понятие файлового сервиса и файлового сервера.
- 2 Интерфейс файлового сервера.
- 3 Интерфейс сервера директорий.
- 4 Семантика разделения файлов.
- 5 Кэширование.
- 6 Поточковая обработка.
- 7 Особенности построения потоковых фреймворков.
- 8 Основные требования к алгоритмам анализа потоковых данных.
- 9 Регионы в Hbase.
- 10 Свойства Hbase.
- 11 Поддерживаемые операции в Hbase.
- 12 Преимущества Cassandra File System (CFS).
- 13 Модель данных в CFS.
- 14 Развертывание и интеграция Cassandra.
- 15 Возможности платформы RapidMine.
- 16 Возможности платформы Statistica Data Miner.
- 17 Преимущества и недостатки нейротехнологии.
- 18 Классификация нейросетевых архитектур.
- 19 Классификация методов обучения нейронных сетей.
- 20 Преимущества и недостатки вероятностных нейронных сетей.
- 21 2Сеть Кохонена.
- 22 Адаптивная нечеткая нейросистема (ANFIS).

#### 4.3 Промежуточная аттестация по дисциплине проводится в форме зачета



## Типовые вопросы зачета

### Типовые вопросы зачета

1. Архитектура распределенных файловых систем.
2. Реализация распределенных файловых систем.
3. Система NFS.

## Типовые задания для зачета

### Типовые задания зачета

- 1 Работа в NFS.
- 2 Работа в GFS.
- 3 Работа в HDFS.

#### 4.4. Шкала оценивания промежуточной аттестации

Оценка	Компетенции	Дескрипторы (уровни) – основные признаки освоения (показатели достижения результата)
«зачтено» (50 - 100 баллов)		
«не зачтено» (0 - 49 баллов)		

## 5. Методические указания для обучающихся по освоению дисциплины (модуля)

### 5.1 Методические указания по организации самостоятельной работы обучающихся:

Приступая к изучению дисциплины, в первую очередь обучающимся необходимо ознакомиться содержанием рабочей программы дисциплины (РПД), которая определяет содержание, объем, а также порядок изучения и преподавания учебной дисциплины, ее раздела, части.

Для самостоятельной работы важное значение имеют разделы «Объем и содержание дисциплины», «Учебно-методическое и информационное обеспечение дисциплины» и «Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы».

В разделе «Объем и содержание дисциплины» указываются все разделы и темы изучаемой дисциплины, а также виды занятий и планируемый объем в академических часах.

В разделе «Учебно-методическое и информационное обеспечение дисциплины» указана рекомендуемая основная и дополнительная литература.

В разделе «Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы» содержится перечень профессиональных баз данных и информационных справочных систем, необходимых для освоения дисциплины.

### 5.2 Рекомендации обучающимся по работе с теоретическими материалами по дисциплине

При изучении и проработке теоретического материала необходимо:

- просмотреть еще раз презентацию лекции в системе MOODLe, повторить законспектированный на лекционном занятии материал и дополнить его с учетом рекомендованной дополнительной литературы;
- при самостоятельном изучении теоретической темы сделать конспект, используя рекомендованные в РПД источники, профессиональные базы данных и информационные справочные системы;
- ответить на вопросы для самостоятельной работы, по теме представленные в пункте 3.2 РПД.
- при подготовке к текущему контролю использовать материалы фонда оценочных средств (ФОС).

### 5.3 Рекомендации по работе с научной и учебной литературой

Работа с основной и дополнительной литературой является главной формой самостоятельной работы и необходима при подготовке к устному опросу на семинарских занятиях, к дебатам, тестированию, экзамену. Она включает проработку лекционного материала и рекомендованных источников и литературы по тематике лекций.

Конспект лекции должен содержать реферативную запись основных вопросов лекции, в том числе с опорой на размещенные в системе MOODLe презентации, основных источников и литературы по темам, выводы по каждому вопросу. Конспект может быть выполнен в рамках распечатки выдачи презентаций лекций или в отдельной тетради по предмету. Он должен быть аккуратным, хорошо читаемым, не содержать не относящуюся к теме информацию или рисунки.

Конспекты научной литературы при самостоятельной подготовке к занятиям должны содержать ответы на каждый поставленный в теме вопрос, иметь ссылку на источник информации с обязательным указанием автора, названия и года издания используемой научной литературы. Конспект может быть опорным (содержать лишь основные ключевые позиции), но при этом позволяющим дать полный ответ по вопросу, может быть подробным. Объем конспекта определяется самим студентом.

В процессе работы с основной и дополнительной литературой студент может:

- делать записи по ходу чтения в виде простого или развернутого плана (создавать перечень основных вопросов, рассмотренных в источнике);
- составлять тезисы (цитирование наиболее важных мест статьи или монографии, короткое изложение основных мыслей автора);
- готовить аннотации (краткое обобщение основных вопросов работы);
- создавать конспекты (развернутые тезисы).

#### 5.4. Рекомендации по подготовке к отдельным заданиям текущего контроля

Собеседование предполагает организацию беседы преподавателя со студентами по вопросам практического занятия с целью более обстоятельного выявления их знаний по определенному разделу, теме, проблеме и т.п. Все члены группы могут участвовать в обсуждении, добавлять информацию, дискутировать, задавать вопросы и т.д.

Устный опрос может применяться в различных формах: фронтальный, индивидуальный, комбинированный. Основные качества устного ответа подлежащего оценке:

- правильность ответа по содержанию;
- полнота и глубина ответа;
- сознательность ответа;
- логика изложения материала;
- рациональность использованных приемов и способов решения поставленной учебной задачи;
- своевременность и эффективность использования наглядных пособий и технических средств при ответе;
- использование дополнительного материала;
- рациональность использования времени, отведенного на задание.

Устный опрос может сопровождаться презентацией, которая подготавливается по одному из вопросов практического занятия. При выступлении с презентацией необходимо обращать внимание на такие моменты как:

- содержание презентации: актуальность темы, полнота ее раскрытия, смысловое содержание, соответствие заявленной темы содержанию, соответствие методическим требованиям (цели, ссылки на ресурсы, соответствие содержания и литературы), практическая направленность, соответствие содержания заявленной форме, адекватность использования технических средств учебным задачам, последовательность и логичность презентуемого материала;
- оформление презентации: объем (оптимальное количество), дизайн (читаемость, наличие и соответствие графики и анимации, звуковое оформление, структурирование информации, соответствие заявленным требованиям), оригинальность оформления, эстетика, использование возможности программной среды, соответствие стандартам оформления;
- личностные качества: ораторские способности, соблюдение регламента, эмоциональность, умение ответить на вопросы, систематизированные, глубокие и полные знания по всем разделам программы;

- содержание выступления: логичность изложения материала, раскрытие темы, доступность изложения, эффективность применения средств ИКТ, способы и условия достижения результативности и эффективности для выполнения задач своей профессиональной или учебной деятельности, доказательность принимаемых решений, умение аргументировать свои заключения, выводы.

## 6. Учебно-методическое и информационное обеспечение дисциплины

### 6.1 Основная литература:

1. Свистунов, А. Н. Построение распределенных систем на Java. - 2021-01-23; Построение распределенных систем на Java. - Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. - 317 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/73707.html>
2. Кудинов, Ю. И. Интеллектуальные информационные системы : учебное пособие для спо. - Весь срок охраны авторского права; Интеллектуальные информационные системы. - Липецк, Саратов: Липецкий государственный технический университет, Профобразование, 2020. - 63 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/92828.html>
3. Чубукова, И. А. Data Mining : учебное пособие. - 2021-12-05; Data Mining. - Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. - 469 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/89404.html>
4. Воронов, В. И., Воронова, Л. И., Усачев, В. А. Data Mining - технологии обработки больших данных : учебное пособие. - 2024-02-26; Data Mining - технологии обработки больших данных. - Москва: Московский технический университет связи и информатики, 2018. - 47 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/81324.html>

### 6.2 Дополнительная литература:

1. Турута, Е. Н. Учебно-методическое пособие по дисциплине Интеллектуальные информационные системы и технологии. - 2022-04-04; Учебно-методическое пособие по дисциплине Интеллектуальные информационные системы и техн. - Москва: Московский технический университет связи и информатики, 2014. - 24 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/61479.html>
2. Громов, Ю. Ю., Иванова, О. Г., Алексеев, В. В., Беляев, М. П., Швец, Д. П., Елисеев, А. И. Интеллектуальные информационные системы и технологии : учебное пособие. - Весь срок охраны авторского права; Интеллектуальные информационные системы и технологии. - Тамбов: Тамбовский государственный технический университет, ЭБС АСВ, 2013. - 244 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/63850.html>
3. Серегин М. Ю., Ивановский М. А., Яковлев А. В. Интеллектуальные информационные системы : учебное пособие. - Тамбов: Тамбовский государственный технический университет (ТГТУ), 2012. - 205 с. - Текст : электронный // ЭБС «Университетская библиотека онлайн» [сайт]. - URL: <http://biblioclub.ru/index.php?page=book&id=277790>
4. Федин, Ф. О., Федин, Ф. Ф. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие. - Весь срок охраны авторского права; Анализ данных. Часть 2. Инструменты Data Mining. - Москва: Московский городской педагогический университет, 2012. - 308 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/26445.html>
5. Волкова, Т. В., Насейкина, Л. Ф. Разработка систем распределенной обработки данных : учебно-методическое пособие. - Весь срок охраны авторского права; Разработка систем распределенной обработки данных. - Оренбург: Оренбургский государственный университет, ЭБС АСВ, 2012. - 330 с. - Текст : электронный // IPR BOOKS [сайт]. - URL: <http://www.iprbookshop.ru/30127.html>

### 6.3 Иные источники:

1. Национальный Открытый Университет «ИНТУИТ» - <http://www.intuit.ru/>

2. СКА MahtPartner - <http://mathpar.cloud.unihub.ru/>

3. <http://hbase.apache.org/book/book.html> - <http://hbase.apache.org/book/book.html>

## **7. Материально-техническое обеспечение дисциплины, программное обеспечение, профессиональные базы данных и информационные справочные системы**

Для проведения занятий по дисциплине необходимо следующее материально-техническое обеспечение: учебные аудитории для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы.

Учебные аудитории и помещения для самостоятельной работы укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещения для самостоятельной работы укомплектованы компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду Университета.

Для проведения занятий лекционного типа используются наборы демонстрационного оборудования, обеспечивающие тематические иллюстрации (проектор, ноутбук, экран/ интерактивная доска).

Лицензионное и свободно распространяемое программное обеспечение:

LibreOffice

Microsoft Office Профессиональный плюс 2007

Microsoft Windows 10

Операционная система "Альт Образование"

Профессиональные базы данных и информационные справочные системы:

1. Springer Journal – база данных журналов коллекции Springer Journal изд-ва Springer Nature (1997-2015 гг.). – URL: <https://link.springer.com>

2. Президентская библиотека имени Б.Н. Ельцина. – URL: <https://www.prilib.ru>

3. Российская государственная библиотека: официальный сайт. – URL: <https://www.rsl.ru>

4. Российская национальная библиотека: официальный сайт. – URL: <http://nlr.ru>

5. Тамбовская областная универсальная научная библиотека им. А.С. Пушкина: официальный сайт. – URL: <http://www.tambovlib.ru>

6. Университетская библиотека онлайн: электронно-библиотечная система. – URL: <https://biblioclub.ru>

7. ЭБС «Университетская библиотека онлайн» . – URL: <http://www.biblioclub.ru>

8. Юрайт: образовательная платформа, электронно-библиотечная система. – URL: <https://urait.ru>

### **Электронная информационно-образовательная среда**

[https://auth.tsutmb.ru/authorize?response\\_type=code&client\\_id=moodle&state=xyz](https://auth.tsutmb.ru/authorize?response_type=code&client_id=moodle&state=xyz)

Взаимодействие преподавателя и студента в процессе обучения осуществляется посредством мультимедийных, гипертекстовых, сетевых, телекоммуникационных технологий, используемых в электронной информационно-образовательной среде университета.