

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт информационных технологий и интеллектуальных систем



подписано электронно-цифровой подписью

Программа дисциплины

Дополнительные главы анализа данных

Направление подготовки: 09.03.04 - Программная инженерия
Профиль подготовки: Цифровая аналитика и инженерия данных
Квалификация выпускника: бакалавр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2025

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): директор института информационных технологий и интеллектуальных систем Абрамский М.М. (Дирекция ИТИС, Институт информационных технологий и интеллектуальных систем), mabramsk@kpfu.ru ; старший преподаватель, б/с Лукьяничева Е.О. (Кафедра программной инженерии, Институт информационных технологий и интеллектуальных систем), EOLukyanicheva@kpfu.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-11	Способность применять перспективные методы исследования и решения профессиональных задач на основе знания мировых тенденций развития технологий искусственного интеллекта

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- Продвинутую статистику: PCA, факторный анализ, многомерные методы
- Методы ML: классификация (SVM, ансамбли), регрессия, отбор признаков, кросс-валидация
- Методы без учителя: кластеризация (K-means, DBSCAN, иерархическая), выявление аномалий
- NLP: обработка текстов, embeddings, трансферное обучение, предобученные модели, LLM
- Временные ряды: разложение, ARIMA, Prophet, LSTM, прогнозирование
- Глубокое обучение: нейросети, CNN, RNN, трансформеры, интерпретация (LIME, SHAP)

Должен уметь:

- Применять продвинутые статистические методы к высокомерным данным
- Разрабатывать модели ML, проводить отбор признаков, оптимизацию гиперпараметров
- Кластеризовать данные, выявлять аномалии, выбирать оптимальное число кластеров
- Обращивать текстовые данные, проводить анализ тональности, работать с LLM
- Анализировать и прогнозировать временные ряды
- Разрабатывать нейросетевые модели, применять методы интерпретации

Должен владеть:

- Практической работой с scikit-learn, XGBoost, LightGBM, CatBoost
- Разработкой нейросетей на TensorFlow/Keras и PyTorch
- Работой с NLP библиотеками (spaCy, NLTK, transformers)
- Анализом временных рядов (statsmodels, prophet)
- Применением методов интерпретации (LIME, SHAP)
- Работой с облачными сервисами (Google Colab, Kaggle)

Должен демонстрировать способность и готовность:

- Решением сложных профессиональных задач через AI методы
- Выбором подходящих методов и моделей на основе задачи
- Работой с современными фреймворками, адаптация к новым технологиям
- Критической оценки результатов, обеспечение интерпретируемости

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.05.01 Дисциплины (модули)" основной профессиональной образовательной программы 09.03.04 "Программная инженерия (Цифровая аналитика и инженерия данных)" и относится к дисциплинам по выбору части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 4 курсе в 7 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 5 зачетных(ые) единиц(ы) на 180 часа(ов).

Контактная работа - 72 часа(ов), в том числе лекции - 0 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 72 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 108 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет с оценкой в 7 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Се-местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само-стоя-тельная ра-бота
			Лекции, всего	Лекции в эл. форме	Практи-ческие занятия, всего	Практи-ческие в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
1.	Тема 1. Продвинутая статистика и многомерный анализ	7	0	0	0	0	12	0	18
2.	Тема 2. Машинное обучение: классификация и регрессия	7	0	0	0	0	12	0	18
3.	Тема 3. Методы кластеризации и выявление аномалий	7	0	0	0	0	12	0	18
4.	Тема 4. Обработка текстов и анализ естественного языка	7	0	0	0	0	12	0	18
5.	Тема 5. Анализ временных рядов и прогнозирование	7	0	0	0	0	12	0	18
6.	Тема 6. Глубокое обучение и интерпретация моделей	7	0	0	0	0	12	0	18
	Итого		0	0	0	0	72	0	108

4.2 Содержание дисциплины (модуля)

Тема 1. Продвинутая статистика и многомерный анализ

Расширение знаний по статистике для работы с высокомерными данными. Методы: анализ главных компонент (PCA) для снижения размерности, факторный анализ для выявления скрытых факторов, дискриминантный анализ для классификации. Изучение методов парных сравнений, ANOVA, многомерных распределений. Обсуждение вызовов: проклятие размерности, мультиколлинеарность, интерпретация. Практические задания: применение методов к датасетам, визуализация многомерных данных, оценка значимости компонент.

Тема 2. Машинное обучение: классификация и регрессия

глубление знания ML моделей: логистическая регрессия с регуляризацией, SVM с различными ядрами, ансамбли (Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost). Отбор признаков, кросс-валидация, работа с дисбалансом классов (SMOTE, class weights). Метрики: precision, recall, F1, ROC-AUC для классификации; MAE, RMSE, R2 для регрессии. Оптимизация гиперпараметров (GridSearch, RandomSearch, Bayesian). Практические задания: разработка моделей, оптимизация, сравнение.

Тема 3. Методы кластеризации и выявление аномалий

Методы обучения без учителя: K-means, иерархическая кластеризация (linkage methods), DBSCAN для произвольных форм, Gaussian Mixture Models (GMM). Выявление аномалий: Isolation Forest, Local Outlier Factor, One-Class SVM. Метрики качества: силуэт, Davies-Bouldin index, Calinski-Harabasz. Применения: сегментация клиентов, обнаружение фрода, анализ данных. Практические задания: кластеризация, выбор оптимального числа кластеров, выявление аномалий.

Тема 4. Обработка текстов и анализ естественного языка

Продвинутые методы работы с текстами: токенизация, stemming, lemmatization. Методы признаков: Bag of Words, TF-IDF, word embeddings (Word2Vec, GloVe, FastText). Трансферное обучение: BERT, GPT, T5. Приложения: классификация текстов, анализ тональности, извлечение информации (NER), machine translation. Большие языковые модели (LLM) и API. Практические задания: классификация, анализ тональности, работа с моделями, custom модели.

Тема 5. Анализ временных рядов и прогнозирование

Компоненты временного ряда: тренд, сезонность, цикличность. Методы разложения: STL, seasonal decompose. Классические подходы: ARIMA, SARIMA, экспоненциальное сглаживание. Современные методы: Prophet, LSTM, трансформеры. Вызовы: нестационарность, пропуски, календарные эффекты. Практические задания: разложение, прогнозирование на различных горизонтах, оценка качества.

Тема 6. Глубокое обучение и интерпретация моделей

Архитектуры: полносвязные сети, CNN, RNN/LSTM/GRU, трансформеры. Фреймворки: TensorFlow/Keras, PyTorch. Техники: dropout, batch normalization, early stopping. Интерпретация: LIME, SHAP, attention weights. Практические задания: разработка нейросетей, работа с предобученными моделями, интерпретация решений, облачные сервисы.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Coursera - <https://www.coursera.org/specializations/machine-learning-data-analysis>

Stepik: Анализ данных - <https://stepik.org/course/76/syllabus>

Открытое образование: Анализ данных - <https://openedu.ru/course/spbu/DATAAN/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лабораторные работы	Лабораторные работы предназначены для практического освоения продвинутых методов анализа данных и AI. Перед началом: тщательно ознакомьтесь с теоретическим материалом (конспект, литература, документация библиотек scikit-learn.org , pytorch.org , tensorflow.org , spacy.io , huggingface.co). Подготовьте окружение: установите pandas, numpy, scikit-learn, TensorFlow, PyTorch, используйте Jupyter Notebook.
самостоятельная работа	Самостоятельная работа включает углублённое изучение, практические задания, проекты. Алгоритм: (1) ознакомьтесь с конспектом, определите ключевые вопросы; (2) прочитайте учебники, статьи, видео, составьте конспект; (3) работайте с примерами, модифицируйте параметры; (4) изучите кейсы из практики; (5) выполните задания на датасетах (Kaggle, UCI ML); (6) разработайте собственный проект.
зачет с оценкой	Подготовка: за 2-3 недели повторите все разделы, составьте конспекты, решите практические задачи. Формат: тест, устный экзамен, защита проекта или комбинация. Демонстрировать: понимание методов, практические навыки, критическое мышление, знание трендов, ясную коммуникацию. Ответ: логичный, аргументированный, практичный, применяющий терминологию.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 09.03.04 "Программная инженерия" и профилю подготовки "Цифровая аналитика и инженерия данных".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.05.01 Дополнительные главы анализа данных

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 09.03.04 - Программная инженерия
Профиль подготовки: Цифровая аналитика и инженерия данных
Квалификация выпускника: бакалавр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2025

Основная литература:

1. Криволапов, С. Я. Введение в анализ данных. Поиск структуры данных с применением языка Python : учебное пособие / С.Я. Криволапов. - Москва : ИНФРА-М, 2024. - 177 с. - (Высшее образование). - DOI 10.12737/2082643. - ISBN 978-5-16-019001-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2141600> (дата обращения: 10.12.2024). - Режим доступа: по подписке.
2. Криволапов, С. Я. Анализ данных. Методы теории вероятностей и математической статистики на языке Python : учебное пособие / С.Я. Криволапов. - Москва : ИНФРА-М, 2025. - 678 с. - (Высшее образование). - DOI 10.12737/2034420. - ISBN 978-5-16-018616-0. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2034420> (дата обращения: 10.12.2024). - Режим доступа: по подписке.
3. Вольфсон, М. Б. Анализ данных : учебно-методическое пособие / М. Б. Вольфсон. - Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2023. - 69 с. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/381533> (дата обращения: 10.12.2024). - Режим доступа: для авториз. пользователей.

Дополнительная литература:

1. Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие / Т. М. Татарникова. - Москва ; Вологда : Инфра-Инженерия, 2024. - 172 с. - ISBN 978-5-9729-1772-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2169704> (дата обращения: 10.12.2024). - Режим доступа: по подписке.
2. Баяк, О. А. Практикум по анализу данных на языках Python и R : учебное пособие / О. А. Баяк, М. Р. Исаева, М. О. Самсонкин. - Москва : Прометей, 2023. - 100 с. - ISBN 978-5-00172-356-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2124862> (дата обращения: 10.12.2024). - Режим доступа: по подписке.
3. Танимура К. SQL для анализа данных: Пер. с англ. / К. Танимура. - Санкт-Петербург : БХВ-Петербург, 2024. - 384 с. - ISBN 978-5-9775-0958-9. - URL: <https://ibooks.ru/bookshelf/396435/reading> (дата обращения: 10.12.2024). - Текст: электронный.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.05.01 Дополнительные главы анализа данных

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Цифровая аналитика и инженерия данных

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows