

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



подписано электронно-цифровой подписью

Программа дисциплины
Машинное обучение в задачах прогнозирования

Направление подготовки: 01.04.02 - Прикладная математика и информатика
Профиль подготовки: Анализ данных и его приложения
Квалификация выпускника: магистр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2024

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): старший преподаватель, к.н. Арабов М.К. (кафедра анализа данных и технологий программирования, отделение фундаментальной информатики и информационных технологий), MKAraбов@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-4	Способен строить математические модели и анализировать данные, обосновывать и выбирать решение в прикладных задачах

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

Основные понятия и методы прогнозирования, включая предобработку данных, линейные и матричные методы, анализ временных рядов и современные алгоритмы машинного обучения, такие как градиентный бустинг, ARIMA, нейронные сети и трансформеры.

Этапы процесса прогнозирования и критерии оценки моделей.

Должен уметь:

Применять методы предобработки данных, включая очистку, обработку пропущенных значений и нормализацию. Строить и оценивать регрессионные модели, а также применять регуляризацию для улучшения качества прогнозов. Анализировать временные ряды и использовать различные методы для их прогнозирования. Настраивать гиперпараметры для алгоритмов градиентного бустинга, ARIMA и нейронных сетей. Реализовывать модели многослойных перцептронов, сверточных и рекуррентных нейронных сетей, а также трансформеров.

Должен владеть:

Навыками работы с современными инструментами и библиотеками для анализа данных и машинного обучения. Умением обосновывать выбор методов и моделей в зависимости от специфики задач.

Должен демонстрировать способность и готовность:

Осуществлять прогнозирование на основе различных моделей и методов, адаптируя подходы к конкретным задачам в области экономики и управления. Применять полученные знания и навыки для принятия обоснованных решений в профессиональной деятельности, включая стратегическое планирование и анализ данных.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.03.01 Дисциплины (модули)" основной профессиональной образовательной программы 01.04.02 "Прикладная математика и информатика (Анализ данных и его приложения)" и относится к дисциплинам по выбору части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2 курсе в 3 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 2 зачетных(ые) единиц(ы) на 72 часа(ов).

Контактная работа - 32 часа(ов), в том числе лекции - 16 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 16 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 40 часа(ов).

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

Форма промежуточного контроля дисциплины: зачет в 3 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- тель- ная ра- бота
			Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабора- торные работы, всего	Лабора- торные в эл. форме	
1.	Тема 1. Тема 1. Введение в прогнозирование. Методы предобработки данных	3	2	0	0	0	2	0	5
2.	Тема 2. Тема 2. Применение линейных и матричных методов в задачах прогнозирования (Регрессионный анализ данных)	3	2	0	0	0	2	0	5
3.	Тема 3. Тема 3. Введение в анализ временных рядов	3	2	0	0	0	2	0	5
4.	Тема 4. Тема 4. Прогнозирование с использованием Градиентного бустинга	3	2	0	0	0	2	0	5
5.	Тема 5. Тема 5. Прогнозирование временных рядов методом семейства ARIMA	3	2	0	0	0	2	0	5
6.	Тема 6. Тема 6. Прогнозирование с помощью многослойных перцептронов (MLP) и сверточных нейронных сетей (CNN)	3	2	0	0	0	2	0	5
7.	Тема 7. Тема 7. Прогнозирование с помощью рекуррентных нейронных сетей	3	2	0	0	0	2	0	5
8.	Тема 8. Тема 8. Прогнозирование с помощью трансформеров	3	2	0	0	0	2	0	5
	Итого		16	0	0	0	16	0	40

4.2 Содержание дисциплины (модуля)

Тема 1. Тема 1. Введение в прогнозирование. Методы предобработки данных

Понятие прогнозирования и его роль в анализе данных.

Основные этапы процесса прогнозирования.

Методы предобработки данных:

Очистка данных.

Обработка пропущенных значений.

Нормализация и стандартизация.

Кодирование категориальных переменных.

Визуализация данных перед анализом.

Выбор признаков и его важность.

Обработка выбросов и аномалий.

Разделение данных на обучающую и тестовую выборки.

Использование библиотек для предобработки данных (например, Pandas, NumPy).

Тема 2. Тема 2. Применение линейных и матричных методов в задачах прогнозирования (Регрессионный анализ данных)

Линейная регрессия: основные принципы и формула.

Множественная линейная регрессия: расширение линейной модели.

Регуляризация (Lasso, Ridge, Elastic Net): зачем и как использовать.

Оценка качества регрессионных моделей: метрики (RMSE, MAE, R²).

Проверка предположений линейной регрессии.
Визуализация регрессионных моделей.
Применение регрессионного анализа в реальных задачах.
Использование библиотек для регрессионного анализа (например, scikit-learn).

Тема 3. Тема 3. Введение в анализ временных рядов

Понятие временного ряда и его характеристики.
Компоненты временного ряда: тренд, сезонность, случайная компонента.
Методы анализа временных рядов: визуализация и декомпозиция.
Автокорреляция и частичная автокорреляция.
Прогнозирование на основе временных рядов: основные подходы.
Модели сглаживания (например, экспоненциальное сглаживание).
Применение анализа временных рядов в бизнесе и экономике.
Использование библиотек для анализа временных рядов (например, statsmodels).

Тема 4. Тема 4. Прогнозирование с использованием Градиентного бустинга

Основы алгоритма градиентного бустинга и его принцип работы.
Применение градиентного бустинга для задач регрессии и классификации.
Настройка гиперпараметров градиентного бустинга.
Сравнение градиентного бустинга с другими методами (например, случайный лес).
Применение градиентного бустинга в реальных задачах.
Визуализация важности признаков в моделях градиентного бустинга.
Проблемы переобучения и методы их решения.
Использование библиотек для градиентного бустинга (например, XGBoost, LightGBM).

Тема 5. Тема 5. Прогнозирование временных рядов методом семейства ARIMA

Модели ARIMA: определение и компоненты.
Идентификация, оценка и диагностика ARIMA-моделей.
Применение ARIMA-моделей для прогнозирования временных рядов.
Модели SARIMA для учета сезонности.
Проверка предположений о стационарности.
Прогнозирование с использованием ARIMA: пошаговый процесс.
Сравнение ARIMA с другими методами прогнозирования.
Использование библиотек для работы с ARIMA (например, statsmodels).

Тема 6. Тема 6. Прогнозирование с помощью многослойных перцептронов (MLP) и сверточных нейронных сетей (CNN)

Архитектура и обучение MLP: основные принципы.
Применение CNN для анализа временных рядов.
Особенности использования нейронных сетей в задачах прогнозирования.
Настройка гиперпараметров нейронных сетей.
Визуализация и интерпретация результатов нейронных сетей.
Проблемы переобучения и методы регуляризации.
Применение MLP и CNN в реальных задачах.
Использование библиотек для нейронных сетей (например, TensorFlow, Keras).

Тема 7. Тема 7. Прогнозирование с помощью рекуррентных нейронных сетей

Принцип работы рекуррентных нейронных сетей (RNN).
Применение RNN (LSTM, GRU) для прогнозирования временных рядов.
Особенности обучения и использования RNN.
Архитектура LSTM и GRU: ключевые отличия.
Проблемы затухающего градиента и их решение.
Применение RNN в задачах обработки последовательностей.
Визуализация результатов работы RNN.
Использование библиотек для RNN (например, TensorFlow, PyTorch).

Тема 8. Тема 8. Прогнозирование с помощью трансформеров

Архитектура трансформеров: ключевые компоненты.
Применение трансформеров для задач прогнозирования.
Преимущества и особенности трансформеров в сравнении с другими моделями.
Обучение трансформеров на временных рядах.
Настройка гиперпараметров трансформеров.
Применение механизмов внимания в трансформерах.
Сравнение трансформеров с RNN и CNN.
Использование библиотек для работы с трансформерами (например, Hugging Face Transformers).

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)

Интернет-портал образовательных ресурсов по ИТ - <http://www.intuit.ru>

Учебник по ML - <https://education.yandex.ru/handbook/ml>

Яндекс образование - <https://education.yandex.ru/journal/avtor-onlayn-uchebnika-po-nlp-o-tom-kak-i-zachem-ego-izuchat>

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

Вид работ	Методические рекомендации
лекции	Студенту рекомендуется внимательно слушать лектора, следить за тем, что написано на доске или представлено на слайдах презентации, анализировать получаемую им информацию. В случае, если материал лекции непонятен, следует задать вопрос в отведенное для вопросов время. Студенту также рекомендуется конспектировать материал лекции в тетради, что улучшает запоминание.
лабораторные работы	При выполнении лабораторных работ студенту рекомендуется внимательно анализировать поставленную задачу, уделяя особенное внимание критериям оценки точности решения задачи. Программный код должен быть объектно-ориентированным, чистым, с поясняющими комментариями. Особенное внимание следует уделить методологическим аспектам решения задачи, на корректное разделение выборки на обучающую, валидационную и тестовую. Результаты работы программы должны быть оформлены в виде таблиц и графиков.
самостоятельная работа	При ведении самостоятельной работы студенту рекомендуется внимательно подходить к изучению научных статей, обращать внимание на значимость полученного результата, на требования к обучающей выборке, на скорость работы предлагаемых алгоритмов, на результаты их сравнения с существующими. В случае, если изучаемый материал понятен не до конца, рекомендуется обращение к дополнительной литературе.
зачет	Студенту рекомендуется внимательно проанализировать вопросы в списке заданий. Ответ на задание должен быть подробным и четким, с приведением и объяснением всех релевантных формул. При ответе на вопрос студент должен продемонстрировать не только знание материала, но и его глубокое понимание. Следует избегать включения в ответ информации, не относящейся к заданию.

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 01.04.02 "Прикладная математика и информатика" и магистерской программе "Анализ данных и его приложения".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.03.01 Машинное обучение в задачах прогнозирования

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Основная литература:

Бокс, Г. Е. П., Дженкинс, Г. М. Временные ряды: анализ и прогнозирование. - М.: Статистика, 1979. - 528 с.

Хан, Д. К. Статистический анализ данных. - М.: Финансы и статистика, 2005. - 384 с.

Маккинни, У. Анализ данных с помощью Python. - М.: Вильямс, 2018. - 400 с.

Гудфеллоу, И., Бенджио, Й., Курвилле, А. Глубокое обучение. - М.: ДМК Пресс, 2019. - 800 с.

Мерфи, К. П. Машинное обучение: вероятностный подход. - М.: ДМК Пресс, 2012. - 800 с.

Сидоров, А. П. Методы предобработки данных. - М.: Наука, 2020. - 256 с.

Шмидт, В. А. Прогнозирование и анализ временных рядов. - М.: Экономика, 2011. - 300 с.

Дополнительная литература:

Куров, В. Н. Основы машинного обучения. - М.: ИНФРА-М, 2017. - 350 с.

Румельт, Р. П. Анализ данных и принятие решений. - М.: ИДБ, 2015. - 450 с.

Петров, И. А. Статистические методы в анализе данных. - М.: Высшая школа, 2013. - 320 с.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.03.01 Машинное обучение в задачах прогнозирования

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

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

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

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.