

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования_
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Институт экономики и предпринимательства

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Интеллектуальный анализ данных и процессов

Уровень высшего образования

Магистратура

Направление подготовки / специальность

38.04.01 - Экономика

Направленность образовательной программы

Информационная безопасность экономических процессов

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.04.02 Интеллектуальный анализ данных и процессов относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-11: Способен управлять ресурсами, процессами, службами и подразделениями, осуществляющими экономическую деятельность на предприятиях и организациях различных форм собственности, в органах государственной и муниципальной власти	ПК-11.1: Управляет ресурсами и процессами на предприятиях и организациях различных форм собственности, в органах государственной и муниципальной власти ПК-11.2: Руководит службами и подразделениями, осуществляющими аналитическую, производственно-хозяйственную, организационно-экономическую и внешнеэкономическую деятельность на предприятиях и организациях различных форм собственности	ПК-11.1: Знать: основные понятия и принципы интеллектуального анализа данных, включая методы машинного обучения, статистический анализ и обработку больших объемов данных. Уметь: проводить предобработку данных, включая очистку, интеграцию, трансформацию и нормализацию данных. Владеть: инструментами и программным обеспечением для анализа данных, а также специализированными библиотеками для машинного обучения и обработки данных. ПК-11.2: Знать: различные методы и техники анализа данных, включая классификацию, кластеризацию, анализ ассоциативных правил, выявление аномалий, а также специфические подходы к анализу временных рядов и текстов. Уметь применять различные методы анализа данных для решения конкретных задач, интерпретировать полученные результаты и делать обоснованные выводы.	Тест	Зачёт: Проектная работа

		Владеть инструментами и программным обеспечением для анализа данных, а также специализированными библиотеками для машинного обучения и обработки данных.		
ПК-12: Способен организовывать сотрудничество с потенциальными партнерами, разрабатывать варианты управленческих решений, обосновывать их выбор на основе проведенного бизнес-анализа, осуществлять контроль и оценку их эффективности	ПК-12.1: Организует сотрудничество с потенциальными партнерами ПК-12.2: Проводит бизнес-анализ деятельности организации и разрабатывает варианты управленческих решений ПК-12.3: Осуществляет контроль реализации и оценку эффективности управленческих решений	<p>ПК-12.1: Знать этапы процесса анализа данных, начиная с предобработки данных до интерпретации результатов и визуализации. Уметь использовать методы машинного обучения как для обучения с учителем и без учителя, так и для глубокого обучения и разработки рекомендательных систем. Владеть умением эффективно коммуницировать результаты анализа данных различной аудитории, включая специалистов в области данных и неспециалистов, с использованием подходящих методов визуализации и презентации информации.</p> <p>ПК-12.2: Знать: применение анализа данных в различных сферах деятельности, таких как бизнес, здравоохранение, социальные сети, промышленность и производство. Уметь визуализировать данные, создавать интерактивные дашборды и презентации для наглядного представления результатов анализа. Владеть навыками критического мышления к решению проблем, способностью оценивать и интерпретировать сложные наборы данных.</p> <p>ПК-12.3: Знать этические и юридические аспекты работы</p>	Тест	Зачёт: Проектная работа

		с данными, включая вопросы конфиденциальности, защиты информации и регулирования в области больших данных. Уметь реализовывать аналитические проекты от идеи до представления результатов, включая разработку проекта, анализ данных, визуализацию и интерпретацию результатов. Владеть навыками аналитического исследования при решении проблем, способностью оценивать и интерпретировать сложные наборы данных.		
ДПК-2: Способен принимать участие в формировании, организовывать и поддерживать выполнение комплекса мер по обеспечению информационной безопасности, управлять процессом их реализации	ДПК-2.1: организует и управляет процессом реализации комплекса мер по обеспечению информационной безопасности	ДПК-2.1: Знать этические и юридические аспекты работы с данными, включая вопросы конфиденциальности, защиты информации и регулирования в области больших данных. Уметь анализировать тексты и временные ряды с использованием специализированных методов и инструментов. Владеть способностью к самостоятельному обучению и развитию в быстро меняющейся области анализа данных.	Тест	Зачёт: Проектная работа

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	2
Часов по учебному плану	72
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	8
- занятия семинарского типа (практические занятия / лабораторные работы)	16
- КСР	1
самостоятельная работа	47

Промежуточная аттестация	0 Зачёт
--------------------------	------------

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Тема 1: Введение в интеллектуальный анализ данных.	10	2	2	4	6
Тема 2: Предобработка данных	10	2	2	4	6
Тема 3: Методы и техники анализа данных.	12	2	2	4	8
Тема 4: Машинное обучение в анализе данных.	14	2	4	6	8
Раздел 5: Интеллектуальный анализ временных рядов и текстов.	9		2	2	7
Раздел 6: Визуализация данных.	8		2	2	6
Тема 7: Применение интеллектуального анализа данных в различных областях	8		2	2	6
Аттестация	0				
КСР	1			1	
Итого	72	8	16	25	47

Содержание разделов и тем дисциплины

Тема 1: Введение в интеллектуальный анализ данных. Основные понятия и определения. История развития и текущее состояние области. Основные этапы процесса анализа данных.

Тема 2: Предобработка данных. Очистка данных. Интеграция данных. Трансформация данных. Нормализация и стандартизация данных.

Тема 3: Методы и техники анализа данных. Классификация и регрессия. Кластеризация. Анализ ассоциативных правил. Выявление аномалий.

Тема 4: Машинное обучение в анализе данных. Обучение с учителем и без учителя. Глубокое обучение. Рекомендательные системы. Нейронные сети и их применение.

Тема 5: Интеллектуальный анализ временных рядов и текстов. Анализ временных рядов. Обработка естественного языка (NLP). Сентимент-анализ. Тематическое моделирование.

Тема 6: Визуализация данных. Основы визуализации данных. Инструменты и технологии визуализации. Интерактивная визуализация и дашборды.

Тема 7: Применение интеллектуального анализа данных в различных областях. Этические и юридические аспекты анализа данных. Конфиденциальность и защита данных. Практические проекты и кейс-стади.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для обеспечения самостоятельной работы обучающихся используются:

- электронный курс "-" (-).
- открытый онлайн-курс МООС "-" (-).

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПК-11:

1. Что такое машинное обучение?

1. Процесс создания программного обеспечения для анализа данных.
2. Использование статистических данных для предсказания будущих событий.
3. Обучение компьютеров обучаться на основе данных без явного программирования.
4. Проектирование и создание роботов.

Правильный ответ: С

2. Какая задача обычно решается с помощью регрессии?

5. Предсказание категории объекта.
6. Предсказание числового значения.
7. Разделение данных на группы.
8. Сокращение количества переменных в данных.

Правильный ответ: В

3. Какой метод используется для группировки аналогичных объектов в данных?

1. Классификация
2. Регрессия
3. Кластеризация
4. Ассоциация

Правильный ответ: С

4. Что такое переобучение в машинном обучении?

1. Недостаточное обучение модели на данных.
2. Модель работает идеально на обучающем наборе данных, но плохо на новых данных.
3. Процесс обучения модели на слишком малом количестве данных.
4. Использование слишком малого количества признаков для обучения модели.

Правильный ответ: В

5. Какая библиотека в Python наиболее часто используется для обработки данных?

1. TensorFlow
2. Keras
3. NumPy
4. PyTorch

Правильный ответ: С

5.1.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПК-12:

1. Какой метод анализа данных наиболее подходит для идентификации групп похожих объектов в наборе данных?

1. Регрессия
2. Классификация
3. Кластеризация
4. Анализ ассоциативных правил

Правильный ответ: С

2. Для чего обычно используется анализ временных рядов?

1. Для выявления аномалий в данных.
2. Для предсказания будущих значений на основе прошлых и текущих данных.
3. Для классификации текстов.
4. Для группировки похожих объектов.

Правильный ответ: В

3. Какой метод используется для выявления необычных данных, которые отличаются от остальных?

1. Анализ ассоциативных правил
2. Выявление аномалий
3. Кластеризация
4. Классификация

Правильный ответ: В

4. Что такое анализ ассоциативных правил?

1. Метод для определения силы взаимосвязи между переменными.
2. Процесс классификации текстов.
3. Метод для выявления часто встречающихся комбинаций объектов или явлений.
4. Техника для прогнозирования временных рядов.

Правильный ответ: С

5. Какая библиотека в Python чаще всего используется для анализа и обработки текстов?

1. NumPy
2. Matplotlib
3. Pandas
4. NLTK

Правильный ответ: D

5.1.3 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ДПК-2:

1. Что включает в себя понятие "этика данных"?

1. Разработка алгоритмов машинного обучения
2. Процесс сбора и анализа данных
3. Учет моральных принципов при работе с данными, в том числе конфиденциальности и справедливости
4. Только юридические аспекты регулирования данных

Правильный ответ: C

2. Какой закон регулирует обработку персональных данных в Европейском Союзе?

1. COPPA
2. FERPA
3. GDPR
4. HIPAA

Правильный ответ: C

3. Что такое анонимизация данных?

1. Процесс сбора данных
2. Процесс удаления или изменения персональной информации, чтобы предотвратить идентификацию субъекта данных
3. Шифрование данных
4. Регулярное обновление данных

Правильный ответ: B

4. Для чего применяются специализированные методы анализа текстов?

1. Для улучшения качества данных
2. Для создания более точных моделей временных рядов
3. Для извлечения значимой информации из текстовых данных
4. Для анонимизации данных

Правильный ответ: C

5. Какой метод чаще всего используется при анализе временных рядов?

1. Логистическая регрессия
2. Кластеризация
3. ARIMA
4. Наивный байесовский классификатор

Правильный ответ: C

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	Количество правильных ответов равно, или не менее 50%
не зачтено	Количество правильных ответов менее 50%

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатор достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым	Продemonстрированы базовые навыки при решении стандартных задач с некоторым и	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и	Продemonстрирован творческий подход к решению нестандартных задач

	ответа		и недочетами	недочетами		недочетов	
--	--------	--	-----------------	------------	--	-----------	--

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Проектная работа) для оценки сформированности компетенции ПК-11

Выполните проектную работу по анализу и прогнозированию экономических показателей на примере компании.

Задание 1. Выполните этап "Предобработка данных".

1. Очистка данных от пропусков и аномалий.
2. Интеграция данных из разных источников (если данные будут предоставлены в нескольких файлах).
3. Трансформация категориальных переменных с помощью техник кодирования (например, One-Hot Encoding).
4. Нормализация числовых переменных для улучшения сходимости моделей машинного обучения.

5.3.2 Типовые задания (оценочное средство - Проектная работа) для оценки сформированности компетенции ПК-12

Задание 2. Выполните этапы "Анализ данных" и "Моделирование".

Анализ данных:

1. Исследование распределения ключевых переменных.
2. Выявление и анализ корреляций между переменными.
3. Проведение статистического анализа для выявления значимых факторов, влияющих на выбранный экономический показатель.

Моделирование:

1. Разделение данных на обучающую и тестовую выборки.
2. Построение нескольких моделей машинного обучения (например, логистическая регрессия, случайный лес, градиентный бустинг) для предсказания оттока клиентов.
3. Оценка качества моделей с использованием соответствующих метрик (точность, полнота, F1-мера, AUC-ROC).

5.3.3 Типовые задания (оценочное средство - Проектная работа) для оценки сформированности компетенции ДПК-2

Задание. Выполните этапы "Интерпретация результатов" и "Презентация проекта".

Интерпретация результатов:

1. Анализ важности признаков для принятия решений.
2. Выводы о том, какие факторы наиболее сильно влияют на выбранный экономический показатель.
3. Рекомендации для компании.

Презентация проекта:

1. Подготовка отчета о проделанной работе, включая описание предобработки данных, анализа, выбора и обоснования выбора моделей, а также основных выводов и рекомендаций.
2. Подготовка презентации результатов проекта для заинтересованных сторон.

Критерии оценивания (оценочное средство - Проектная работа)

Оценка	Критерии оценивания
зачтено	Задания выполнены в соответствии с предложенными шаблонами и инструкциями.
не зачтено	Задания не выполнены в соответствии с предложенными шаблонами и инструкциями.

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Тюрин Ю. Н. Анализ данных на компьютере : учебное пособие . - 4-е изд., перераб. - М. : Форум, 2010. - 368 с. - ISBN 978-5-8199-0356-8 : 139.41., 5 экз.
2. Васильев Е. П. Интеллектуальный анализ данных в технологиях принятия решений : учебное пособие / Васильев Е. П., Орешков В. И. - Рязань : РГРТУ, 2023. - 180 с. - Книга из коллекции РГРТУ - Информатика. - ISBN 978-5-7722-0344-6., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=888041&idb=0>.

3. Вольфсон М. Б. Анализ данных : учебно-методическое пособие / Вольфсон М. Б. - Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2023. - 69 с. - Книга из коллекции СПбГУТ им. М.А. Бонч-Бруевича - Информатика., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=888128&idb=0>.
4. Анализ данных : учебник / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. - Москва : Юрайт, 2023. - 490 с. - (Высшее образование). - ISBN 978-5-534-00616-2. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=847079&idb=0>.
5. Целых Александр Николаевич. Современные методы прикладной информатики в задачах анализа данных : Учебное пособие / Южный федеральный университет. - Ростов-на-Дону : Издательство Южного федерального университета (ЮФУ), 2021. - 130 с. - ВО - Бакалавриат. - ISBN 978-5-9275-3783-9., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=835201&idb=0>.
6. Целых Александр Николаевич (Южный федеральный университет). Принятие решений на основе методов машинного обучения : Учебное пособие по курсам «Модели и методы инженерии знаний», «Методы анализа больших данных»; Учебное пособие / Южный федеральный университет; Ростовский государственный экономический университет (РИНХ). - Ростов-на-Дону : Издательство Южного федерального университета (ЮФУ), 2022. - 113 с. - ВО - Бакалавриат. - ISBN 978-5-9275-4246-8., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=890799&idb=0>.
7. Полковникова Наталья Анатольевна. Анализ и визуализация данных в Microsoft Excel в примерах и задачах : Практическое пособие. - Вологда : Инфра-Инженерия, 2023. - 172 с. - Профессиональное образование. - ISBN 978-5-9729-1485-2., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=876044&idb=0>.

Дополнительная литература:

1. Демидова Л. А. Разведочный анализ данных. Python. Ч. 1. Разведочный анализ данных. Python. Часть 1 / Демидова Л. А. - Москва : РТУ МИРЭА, 2022. - 107 с. - Книга из коллекции РТУ МИРЭА - Информатика., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=861303&idb=0>.
2. Миркин Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. - Москва : Юрайт, 2023. - 174 с. - (Высшее образование). - ISBN 978-5-9916-5009-0. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=843081&idb=0>.
3. Демидова Л. А. Разведочный анализ данных. Python. Ч. 2. Разведочный анализ данных. Python. Часть 2 / Демидова Л. А. - Москва : РТУ МИРЭА, 2023. - 92 с. - Книга из коллекции РТУ МИРЭА - Информатика. - ISBN 978-5-7339-1933-1., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=888509&idb=0>.
4. Принятие решений на основе анализа слабоструктурированных данных социальных сетей / Ярушкина Н. Г., Андреев И. А., Желепов А. С., Мошкин В. С., Ярушкина Н. Г. - Ульяновск : УлГТУ, 2022. - 113 с. - Книга из коллекции УлГТУ - Информатика. - ISBN 978-5-9795-2275-3., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=863560&idb=0>.
5. Шихова О. А. Анализ данных в экономике с использованием возможностей MS Excel : учебно-методическое пособие для студентов направления подготовки 38.04.01 экономика / Шихова О. А. - Вологда : ВГМХА им. Н.В. Верещагина, 2022. - 101 с. - Книга из коллекции ВГМХА им. Н.В. Верещагина - Экономика и менеджмент., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=862035&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. Язык программирования: Python
2. Библиотеки для анализа данных: pandas, numpy
3. Библиотеки для визуализации данных: matplotlib, seaborn
4. Библиотеки для машинного обучения

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 38.04.01 - Экономика.

Автор(ы): Чепьюк Ольга Ростиславовна, доктор философских наук, доцент.

Программа одобрена на заседании методической комиссии от 12.12.2023, протокол № 6.