

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Ректор Белорусского
государственного университета

А.Д.Король

27 июня 2025 г.

Регистрационный № 3612/б.



ТЕХНОЛОГИИ АНАЛИЗА И ВИЗУАЛИЗАЦИИ ДАННЫХ

Учебная программа учреждения образования по учебной дисциплине для
специальности:

6-05-0533-11 Прикладная информатика

Профилязация: Программное обеспечение информационных систем

2025 г.

ПРИЛОЖЕНИЯ К АДДИЦИОННОМУ ПРИЛОЖЕНИЮ

Учебная программа составлена на основе ОСВО 6-05-0533-11-2023 и учебных планов БГУ № 6-5.3-59/04 от 15.05.2023 и № 6-5.3-59/12ин. от 31.05.2023.

СОСТАВИТЕЛЬ:

М.М.Лукашевич, доцент кафедры информационных систем управления факультета прикладной математики и информатики Белорусского государственного университета, кандидат технических наук, доцент

РЕЦЕНЗЕНТ:

Б.В.Никульшин, заведующий кафедрой электронных вычислительных машин факультета компьютерных систем и сетей учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», кандидат технических наук, доцент

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой информационных систем управления БГУ
(протокол № 15 от 19.06.2025)

Научно-методическим Советом БГУ
(протокол № 11 от 26.06.2025)

Заведующий кафедрой

М.Лукашевич

А.М.Недзьведь

Ф. Сидорова Т.В.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель преподавания учебной дисциплины «Технологии анализа и визуализации данных» – сформировать у студента теоретические и практические компетенции в области целостного представления, понимания места и роли, применения технологий анализа данных, ознакомить студентов с основным принципом визуализации разных типов данных, дать практические навыки визуализации и интерпретации данных. При изложении дисциплины важно показать возможности использования интегрированных сред разработки и пакетов прикладных программ для решения прикладных задач, возникающих в различных областях науки, техники, экономики и производства.

Основной задачей, решаемой при изучении учебной дисциплины «Технологии анализа и визуализации данных», является подготовка специалиста, владеющего знаниями в области прикладных технологий анализа и визуализации данных; понимающего концепции и технологии современного анализа и визуализации данных; владеющего умениями и навыками самостоятельного решения задач анализа и визуализации данных с использованием современных программных средств.

Учебная дисциплина «Технологии анализа и визуализации данных» направлена на формирование у студентов компетенций в области обработки, интерпретации и визуализации данных. В рамках дисциплины рассматриваются принципы восприятия визуальной информации человеком, а также подходы к выбору эффективных способов графического отображения информации. Особое внимание уделяется методам выявления скрытых закономерностей в данных, включая задачи группировки и разбиения на категории с применением как строгих алгоритмических, так и вероятностных подходов. Изучаются техники упрощения сложных многомерных наборов данных с сохранением их существенных характеристик. В процессе обучения студенты осваивают современные программные инструменты и интегрированные среды, поддерживающие полный цикл аналитической работы — от подготовки и обработки данных до их интерактивной визуализации и интерпретации результатов.

Основой для обучения являются компетенции, сформированные при изучении дисциплин «Алгоритмы и структуры данных», «Основы и методологии программирования».

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина относится к дисциплинам профилизации «Программное обеспечение информационных систем» компонента учреждения образования.

Программа составлена с учётом межпредметных **связей** с учебными дисциплинами. Дисциплина «Технологии анализа и визуализации данных» непосредственно связана с дисциплинами:

– «Дискретная математика и математическая логика», «Методы оптимизации» «Алгоритмы и структуры данных», «Основы и методологии программирования». Теоретические основы, излагаемые в указанных дисциплинах, используются при реализации и оценке эффективности алгоритмов анализа данных, исследовательском анализе данных.

Сформированные при изучении дисциплины «Технологии анализа и визуализации данных» компетенции являются основой для изучения дисциплины «Искусственный интеллект».

Знания, полученные в учебной дисциплине, используются при изучении дисциплин специализации, при выполнении курсовых и дипломных работ, а также используются как инструментарий для моделирования и компьютерного решения задач ряда математических дисциплин, изучаемых на старших курсах.

Требования к компетенциям

Освоение учебной дисциплины «Технологии анализа и визуализации данных» должно обеспечить формирование следующей компетенции:

Универсальные компетенции:

Решать стандартные задачи профессиональной деятельности на основе применения информационно-коммуникационных технологий.

Специализированные компетенции:

Разрабатывать программы для обработки и анализа данных с заданными функциями; внедрять методы анализа и визуализации информации, а также проводить оценку производительности и точности используемых алгоритмов.

В результате освоения учебной дисциплины студент должен:

знать:

- методы и технологии разведочного анализа данных;
- методы и технологии кластерного анализа и классификации данных;
- базовые программно-технические решения анализа данных;
- методы и технологии визуализации данных;
- базовые программно-технические решения визуализации данных;

уметь:

- программировать задачи анализа данных средней сложности;
- оценивать эффективность различных алгоритмических решений анализа данных;
- разрабатывать программные приложения анализа данных с заданной функциональностью;
- реализовывать задачи визуализации данных средней сложности;
- оценивать эффективность различных решений визуализации данных;
- разрабатывать программные приложения визуализации данных с заданной функциональностью;

иметь навык:

- разработки программных модулей на основе прикладных пакетов для анализа данных;
- программирования анализа данных,

- работы с программными модулями и средствами для визуализации данных;
- визуализации данных.

Структура учебной дисциплины

Дисциплина изучается в 5 семестре. В соответствии с учебным планом всего на изучение учебной дисциплины «Технологии анализа и визуализации данных» отведено **для очной формы** получения высшего образования – 108 часов, в том числе 68 аудиторных часов, лекции – 34 часов, лабораторные занятия – 34 часов. **Из них:**

Лекции – 34 часов, лабораторные занятия – 30 часов, управляемая самостоятельная работа (УСР) – 4 часа.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – зачёт.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Технологии визуализации и разведочного анализа данных

Тема 1.1. Основные понятия и определения анализа и визуализации данных

От баз данных к анализу данных. Определение анализа данных. Основные понятия. Типы задач: регрессия, классификация, кластеризация, поиск аномалий, снижение размерностей. История визуализации данных. Основные концепции визуального восприятия графиков. Разбор типов данных и выбор подходящих под них графиков. Работа с количественными и качественными данными.

Тема 1.2. Интегрированные среды разработки и пакеты программ

Библиотеки научных вычислений: NumPy, SciPy, Scikit-learn. Библиотека для работы с табличными данными Pandas. Среда интерактивных вычислений Jupyter Notebook: настройка и установка, основные принципы работы.

Онлайн и оффлайн сервисы визуализации данных. Библиотеки визуализации: Matplotlib, Seaborn. Применение pandas в задаче исследования данных. Виды графиков и диаграмм. Основные элементы диаграммы. Создание диаграммы. Форматы изображений. График функции. Гистограмма.

Тема 1.3. Основные этапы разведочного анализа данных

Загрузка и предварительный осмотр данных. Проверка структуры данных: количество наблюдений, признаков, типы переменных. Выявление пропущенных значений. Обнаружение дубликатов. Анализ распределений количественных признаков. Анализ частотных распределений категориальных признаков. Вычисление основных статистик: среднее, медиана, стандартное отклонение, квартили, минимум, максимум. Обнаружение выбросов. Визуализация распределений: гистограммы, boxplot, плотностные графики. Анализ взаимосвязей между признаками. Визуализация корреляций: тепловые карты, scatter-диаграммы. Сравнение групп по целевой переменной. Формулировка гипотез на основе наблюдений. Подготовка данных к дальнейшему анализу на основе выводов разведочного анализа.

Раздел 2. Технологии и алгоритмы анализа данных

Тема 2.1. Задачи классификации и регрессии

Математическая постановка задач регрессии и классификации. Метрические методы регрессии и классификации: метод ближайших соседей, взвешенный метод ближайших соседей. Метод опорных векторов в задаче классификации. Метод опорных векторов для задачи регрессии. Деревья решений для задачи классификации. Алгоритмы построения деревьев. Ансамбли моделей. Случайный лес.

Тема 2.2. Кластерный анализ

Математическая модель кластеризации. Меры однородности объектов. Расстояние между объектами. Меры близости между кластерами. Методы

кластерного анализа. Иерархические агломеративные методы. Последовательные кластер-процедуры. Метод К-средних. Графическое представление результатов кластер-анализа.

Тема 2.3. Работа с признаками

Предобработка признаков, обработка пропущенных значений, масштабирование и нормализация числовых признаков, преобразование распределений признаков, категориальные признаки, label-кодирование, one-hot-кодирование, частотное кодирование, целевое кодирование, порядковые категориальные переменные, отбор признаков, визуализация важности признаков.

Тема 2.4. Метрики качества моделей. Выбор модели

Метрики в задаче регрессии: MAE, MSE, MAPE, R². Метрики в задаче классификации: кросс-энтропия, precision, recall, F-мера, ROC-кривая, AUC ROC. Обобщающая способность модели и ее оценка: отложенная выборка, кросс-валидация. Визуализация моделей классификации, регрессии, кластерного анализа, результатов понижения размерности. Визуальное представление оценок качества моделей классификации, регрессии, кластерного анализа, результатов понижения размерности.

Тема 2.5. Понижение размерности данных

Сущность главных компонент (PCA), их свойства, геометрическая интерпретация. Графическое представление результатов. Аппроксимация с использованием главных компонент. t-SNE (t-distributed Stochastic Neighbor Embedding), принцип работы, визуализация кластеров с помощью t-SNE. Сравнение PCA и t-SNE по целям и результатам. Применение t-SNE для разведочного анализа и интерпретации высокоразмерных данных.

Тема 2.6. Прикладные задачи анализа и визуализации данных

Основные области: компьютерное зрение, обработка естественного языка, рекомендательные системы, анализ временных рядов, обучение ранжированию, построение выводов по данным. Визуализация геопространственных данных. Практические примеры.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
Очная (дневная) форма получения высшего образования с применением
дистанционных образовательных технологий (ДОТ)

| Номер раздела, темы | Название раздела, темы | Количество аудиторных часов | | | | | Количество часов УСР | Форма контроля |
|---------------------|--|-----------------------------|----------------------|---------------------|----------------------|------|----------------------|---------------------|
| | | Лекции | Практические занятия | Семинарские занятия | Лабораторные занятия | Иное | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| | Технологии анализа и визуализации данных | 34 | | | 30 | | 4 | |
| 1. | Технологии визуализации и разведочного анализа данных | 12 | | | 12 | | | |
| 1.1. | Основные понятия и определения анализа и визуализации данных | 2 | | | | | | Экспресс-опрос |
| 1.2. | Интегрированные среды разработки и пакеты программ | 6 | | | 6 | | | Экспресс-опрос |
| 1.3. | Основные этапы разведочного анализа данных | 4 | | | 6 | | | Экспресс-опрос |
| 2. | Технологии и алгоритмы анализа данных | 22 | | | 18 | | 4 | Экспресс-опрос |
| 2.1. | Задачи классификации и регрессии | 4 | | | 4 | | | РГР |
| 2.2. | Кластерный анализ данных | 4 | | | 4 | | | Экспресс-опрос |
| 2.3. | Работа с признаками | 4 | | | 4 | | | РГР |
| 2.4. | Метрики качества моделей. Выбор модели | 4 | | | | | 4 | экспресс-опрос, РГР |
| 2.5. | Понижение размерности данных | 2 | | | 2 | | | Экспресс-опрос |

| | | | | | | | | | |
|------|---|---|--|--|--|---|--|--|------|
| 2.6. | Прикладные задачи анализа и визуализации данных | 4 | | | | 4 | | | PGRP |
|------|---|---|--|--|--|---|--|--|------|

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература

1. Вандер Плас, Дж. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас ; [пер. с англ. И. Пальти]. - Санкт-Петербург ; Москва ; Минск : Питер, 2023. - 573 с. - URL: <https://ibooks.ru/reading.php?short=1&productid=376830>.
2. Брюс, П. Практическая статистика для специалистов Data Science : 50+ важнейших понятий с использованием R и Python / П. Брюс, Э. Брюс, П. Гедек ; [пер. с англ. А. Логунова]. - 2-е изд., перераб. и доп. - Санкт-Петербург : БХВ-Петербург, 2021. - 346 с. - URL: <https://ibooks.ru/reading.php?short=1&productid=380029>.
3. Уилке, К. Основы визуализации данных : пособие по эффективной и убедительной подаче информации / Клаус Уилке ; [перевод с английского М.А. Райтмана]. - Москва : Эксмо, 2024. - 352 с.
4. Богачев, А. А. Графики, которые убеждают всех : [руководство по визуализации данных] / Александр Богачев. - 2-е изд. - Москва : АСТ, 2024. - 237 с.
5. Сандерсон, С. Excel с Python и R : раскройте потенциал расширенной обработки и визуализации данных / С. Сандерсон, Д. Кун ; [науч. ред. Хесус Мартин де ла Сьерра Сильва и др.] ; пер. с англ. С. Черников ; науч. ред. Д. Квист. - Астана : Спринг Бук, 2025. - 319 с.

Дополнительная литература

1. Devpractice Team. Python. Визуализация данных. Matplotlib. Seaborn. Mayavi. - devpractice.ru. 2020. - 412 с.: ил.
2. Груздев А. В. Предварительная подготовка данных в Python: Том 1. Инструменты и валидация. – М.: ДМК Пресс, 2023. – 816 с.: ил.
3. Груздев А. В. Предварительная подготовка данных в Python. Том 2: План, примеры и метрики качества. – М.: ДМК Пресс, 2023. – 814 с.: ил.
4. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. – 2-е изд., испр. и доп. – Москва : ДМК Пресс, 2020. – 540 с.
5. Грас Д. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2021. - 416 с.

Электронные ресурсы

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. [Электронный ресурс]. – Режим доступа: <http://www.machinelearning.ru>

2. Онлайн средство визуализации данных. [Электронный ресурс]. – Режим доступа: <https://www.gapminder.org/>

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Для диагностики компетенций в рамках учебной дисциплины рекомендуется использовать следующие формы:

1. *Устная форма*: экспресс-опрос.
2. *Устно-письменная форма*: расчетно-графические работы.

В качестве рекомендуемых технических средств диагностики используется обучение, организованное на платформе Moodle (<https://edufpmi.bsu.by>).

Формой промежуточной аттестации по дисциплине учебным планом предусмотрен зачёт.

Примерный перечень заданий для управляемой самостоятельной работы

Тема 2.4. Метрики качества моделей. Выбор модели. (4 ч)

Метрики задач классификации, регрессии и кластеризации. Выбор метрики под задачу. Выбор модели на основе метрики.

(Форма контроля – экспресс-опрос, РГР).

Рекомендуемая тематика расчетно-графических работ

Тема 2.1. Задачи классификации и регрессии.

Тема 2.3. Работа с признаками.

Тема 2.4. Метрики качества моделей. Выбор модели.

Тема 2.6. Прикладные задачи анализа и визуализации данных.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используются следующие методы:

– *метод группового обучения*, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

В качестве технических средств для организации работы в рамках учебной дисциплины рекомендуется использовать Образовательный портал БГУ (<https://edufpmi.bsu.by>) – инструмент с эффективной функциональностью контроля, тренинга и самостоятельной работы.

– *практико-ориентированный подход*, который предполагает освоение содержания образования через решения практических задач; приобретение навыков эффективного выполнения разных видов профессиональной

деятельности; ориентацию на генерирование идей, реализацию групповых студенческих проектов; использование процедур, способов оценивания, фиксирующих профессиональные компетенции.

Примерный перечень вопросов к зачёту

1. Определение анализа данных.
2. Типы задач: обучение с учителем (регрессия, классификация).
3. Типы задач: обучение без учителя (кластеризация, поиск аномалий, снижение размерностей).
4. История визуализации данных.
5. Основные концепции визуального восприятия графиков.
6. Разбор типов данных и выбор подходящих под них графиков.
7. Работа с количественными и качественными данными.
8. Среда интерактивных вычислений Jupyter Notebook: настройка и установка, основные принципы работы.
9. Библиотеки научных вычислений: NumPy, SciPy, Scikit-learn.
10. Библиотека Pandas: объекты Series и DataFrame.
11. Загрузка и выгрузка данных. Обработка пропусков и дубликатов.
12. Применение Pandas в исследовательском анализе данных.
13. Библиотеки визуализации: Matplotlib, Seaborn, Plotly.
14. Виды графиков и диаграмм. Основные элементы диаграммы.
15. Создание графика функции и гистограммы.
16. Основные этапы разведочного анализа данных.
17. Анализ распределений и выбросов. Визуализация: гистограммы, boxplot, плотностные графики.
18. Анализ взаимосвязей между признаками. Тепловые карты корреляций, scatter-диаграммы.
19. Метод ближайших соседей, взвешенный метод ближайших соседей.
20. Метод опорных векторов в задачах классификации и регрессии.
21. Деревья решений для задачи классификации. Алгоритмы построения деревьев.
22. Ансамбли моделей. Случайный лес.
23. Меры однородности объектов. Расстояния между объектами. Метод K-средних.
24. Предобработка признаков: масштабирование, кодирование категориальных признаков.
25. Отбор признаков. Визуализация важности признаков.
26. Метрики в задаче регрессии: MAE, MSE, MAPE, R².
27. Метрики в задаче классификации: precision, recall, F-мера, ROC-кривая, AUC ROC.
28. Обобщающая способность модели: отложенная выборка, кросс-валидация.
29. Сущность главных компонент (PCA), их свойства, геометрическая интерпретация. t-SNE.

30. Визуализация моделей и оценок качества: классификация, регрессия, кластеризация, понижение размерности, геоданные.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

| Название учебной дисциплины, с которой требуется согласование | Название кафедры | Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине | Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола) |
|---|------------------|---|---|
| Учебная дисциплина не требует согласования | | | |

Заведующий кафедрой
информационных систем управления
д.т.н., доцент

А.М.Недзьведь

[19.06.2025]

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УО

на _____ / _____ учебный год

| № п/п | Дополнения и изменения | Основание |
|------------------|-------------------------------|------------------|
| | | |

Учебная программа пересмотрена и одобрена на заседании кафедры информационных систем управления (протокол № _____ от _____ 20_ г.)

Заведующий кафедрой

(степень, звание)

(подпись)

(И.О.Фамилия)

УТВЕРЖДАЮ
Декан факультета

(степень, звание)

(подпись)

(И.О.Фамилия)