

**Можсей Н.П.
Шинкевич Е.А.**
БГУИР, г. Минск

Применение цифровых технологий в преподавании дисциплины «Компьютерный анализ данных»

В работе описываются возможности использования цифровых технологий в преподавании дисциплины «Компьютерный анализ данных» для студентов второй ступени обучения Белорусского государственного университета информатики и радиоэлектроники. Приводятся примеры применения различных программных продуктов, а также алгоритмов машинного обучения.

**Mozhey N.P.
Shinkevich E.A.**
BSUIR
Minsk, Belarus

Application of digital technologies in teaching the discipline «Computer data analysis»

The article discusses the possibilities of using digital technologies in teaching the discipline «Computer data analysis» for second-level students of the Belarusian State University of Informatics and Radioelectronics. Examples of using various software products, as well as machine learning algorithms, are given.

В Беларуси, как и во всем мире, в настоящее время в образовательном пространстве достаточно широко применяются цифровые технологии. Есть несколько определений этого понятия. Мы будем опираться на следующее: это «технологии, которые позволяют создавать, хранить, обрабатывать и распространять данные» [1]. Инструменты искусственного интеллекта также внедряются в образовательную среду, их применение позволяет упростить решение некоторых типичных для преподавателя вопросов, например, ускоряет подбор материала по заданной теме, формирование разнообразных заданий и их проверку. В последнее время все чаще используются в различных сферах деятельности алгоритмы машинного обучения. «Машинное обучение (англ. machine learning) – класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение за счёт применения решений множества сходных задач» [2].

Для преподавания дисциплин естественнонаучного направления цифровые технологии предоставляют очень широкие возможности. Здесь можно упомянуть удобство представления и демонстрации лекционного материала, подготовку большого (по сравнению с классическими лекциями) объема учеб-

ного материала, выполнение практических и лабораторных работ с использованием соответствующих программных продуктов, проведение различных уровней контролируемых мероприятий. На примере одной из задач дисциплины «Компьютерный анализ данных» рассмотрим, как продемонстрировать обучающимся широкие возможности, которые открываются при правильном применении цифровых технологий. Стоит упомянуть, что, как следует из названия дисциплины, все виды занятий проводятся в компьютерном классе с использованием соответствующих программных продуктов, в настоящее время особой популярностью пользуется язык программирования Python – это высокоуровневый язык программирования, который отличается низким порогом вхождения и широкими возможностями.

Работа с любыми данными начинается с их изучения: наличие/отсутствие ошибок, аномальных значений, пропусков, визуализация данных, определение возможного закона распределения, расчет числовых характеристик. Python обладает разнообразными вариантами для решения всех вышеперечисленных задач. Так, например, для визуализации данных можно использовать диаграммы, графики, тепловые карты и т.д. Для определения возможного закона распределения достаточно просто строится гистограмма с наложенным на нее графиком плотности распределения изучаемой случайной величины. Если же нужно исследовать зависимость изменения вида графика плотности от изменения параметров изучаемой величины, то и здесь есть соответствующие функции (см. рис. 1).

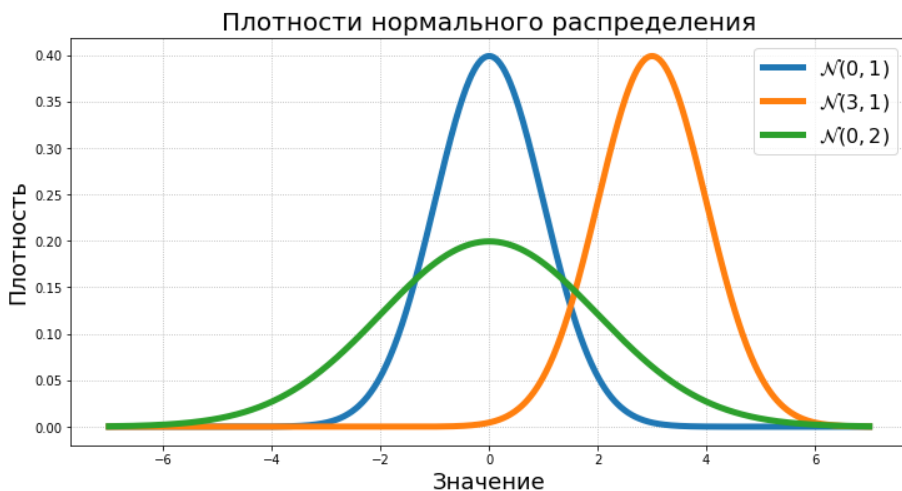


Рис. 1 – Зависимость плотности от изменения параметров

Также следует отметить, что в Python существует такое понятие, как DataFrame – структуры, состоящие из данных, организованных в двух измерениях. Преобразовав массив данных в такую структуру, можно выполнять опе-

рации как со всем массивом данных, так и по столбцам или строкам. Так, например, используя соответствующую библиотеку, можно получить числовые характеристики сразу по всем необходимым столбцам.

После изучения данных и подготовки их к обработке можно проводить анализ. Рассмотрим одну из типичных задач анализа данных: расчет коэффициентов уравнения линейной регрессии. Здесь существует выбор способа реализации. Например, можно использовать классическую схему – составить систему нормальных уравнений и найти коэффициенты уравнения регрессии, решив указанную систему, затем оценить качество полученного уравнения. Можно рассмотреть определение коэффициентов уравнения регрессии как задачу нелинейной оптимизации и использовать для ее решения какой-либо поисковый метод или соответствующий программный продукт. Можно применить готовую библиотеку, например, `scikit-learn`, и также получить коэффициенты линейного уравнения регрессии. Можно воспользоваться готовыми пакетами как для решения оптимизационных задач, так и для построения уравнения регрессии, соответствующие возможности есть даже в Excel или, например, в MathCad.

А еще можно использовать алгоритмы машинного обучения Python. С точки зрения машинного обучения данная задача относится к классу задач обучения с учителем, т.е. модель учится на размеченных наборах данных и сопоставляет точки данных с наиболее подходящими линейными функциями, которые можно использовать для прогнозирования новых данных. Разумеется, какой бы способ не был использован, требования к исходным данным для построения уравнения регрессии едины: данные должны удовлетворять условиям Гаусса-Маркова для регрессионных моделей. Студент освобождается от проведения громоздких вычислений, у него появляется возможность научиться делать выводы на основании полученных данных, определять по результатам возможное нарушение классических предположений, при необходимости изменять полученную модель, чтобы ее можно было использовать для практических целей.

Применение цифровых технологий для сбора и обработки информации в сочетании с соответствующим программным обеспечением позволяет автоматизировать «техническую» сторону решения, существенно увеличить объем расчетов, ускорить оценку и отбор различных вариантов решений. Внедрение современных информационных технологий в образовательный процесс дает возможность повысить качество обучения и подготовить студентов к жизни и работе в условиях цифровой экономики.

Литература:

1. Почему цифровые технологии вытесняют аналоговые. - URL: <https://trends.rbc.ru/trends/industry/60e427ea9a79471089a0ec1d?from=co> (дата обращения: 16.02.2025)
2. Флах П. Машинное обучение. М.: ДМК Пресс, 2015. 400 с.