

*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
им. В.И. Ульянова (Ленина,) г. Санкт-Петербург, Россия*

***Аннотация.** В работе рассматривается опыт разработки учебно-методических материалов для дисциплины «Данные и визуальная аналитика». Представляются возможные темы практических работ, детально рассматривается практическая работа, посвященная разработке методики визуализации многомерных данных. Обсуждается структура практической работы и обосновывается выбор библиотеки визуализации данных.*

Ключевые слова: практические занятия; методические указания; задания; магистерская программа; визуализация данных; лица Чернова

Визуализация данных является важным способом исследования данных, который имеет важное значения на всех этапах анализа данных. Применение методик визуализации позволяет корректно интерпретировать данные и объяснять получаемые результаты применения методов машинного обучения.

Цикл практических работ по курсу «Данные и визуальная аналитика» строится таким образом, чтобы студенты постепенно погружались в задачи визуализации данных, осваивали библиотеки визуализации и учились обоснованно выбирать методики графического представления данных. В настоящее время цикл практических работ состоит из четырех практических работ. Первая практическая работа, названная «Анализ и критика модели визуализации данных» знакомит студентов с основными принципами проектирования методик визуализации – принципами эффективности и выразительности, им предлагается проанализировать различные графики, представленные в научной литературе и СМИ, выявить в них недостатки, обосновать их и разработать свой вариант визуализации данных. Вторая практическая работа «Визуализация многомерных данных. Введение в Matplotlib» знакомит учащихся со стандартными моделями визуализации и показывает, как они могут быть использованы для исследования данных. В ходе выполнения работы студенты осваивают также базовые приемы работы с библиотекой визуализации данных Matplotlib. Третья практическая работа посвящена использованию и разработке нестандартных моделей визуализации многомерных данных для их исследования. Студенты продолжают осваивать библиотеку Matplotlib и учатся создавать собственные модели визуализации. Четвертая практическая работа знакомит с базовыми методиками визуализации графов – диаграммами вида «узел-связь» и матрицей смежности. Студентам предлагается выявить и представить графические различные структуры в графе.

Освоение проблем визуализации данных в задачах обработки и интерпретации данных принадлежит в актуальным задачам формирования глобального информационного общества. Для того, чтобы сделать практические работы более привлекательными для студентов, при их проектировании используются актуальные наборы данных с таких площадок как Kaggle, IEEEDataport, а в качестве методик визуализации предлагается использовать такие способы представления данных, применение которых позволит создавать диаграммы, которые не только полезны, но и интересны. Например, одним из вариантов визуализации многомерных данных являются лица Чернова. Они были разработаны американским математиком Германом Черновым в 1973 году, результаты были изложены в соответствующей научной работе [1]. Согласно результатам исследования, использование лиц в качестве пиктограмм очень эффективно с точки зрения их восприятия человеком в виду того, что люди социальны и способны быстро распознавать мимику и ее динамику. В эту технику визуализации изначально было заложено до восемнадцати параметров, которые включали в себя позиции и размеры черт лица, позднее их количество было увеличено до 30. Стоит отметить, что некоторые черты лица воспринимаются человеком быстрее и точнее остальных, поэтому самые важные параметры данных рекомендуется связывать именно с ними. Лица Чернова чаще всего применяются в

двух случаях: для выявления взаимосвязей в многомерных данных и для выявления сложных зависимостей между переменными в этих данных [2]. Стоит также отметить и недостатки лиц Чернова как техники визуализации – в некоторых случаях восприятие лиц может быть субъективно, кроме того лица Чернова хоть и используются для визуализации данных из разных предметных областей, но все же не имеют с ними контекстуальной связи, то есть существует возможность разработки похожей концепции визуализации, но с учетом контекста предметной области, что также снимет некоторые вопросы о субъективности восприятия результатов.

Количество параметров и визуализация лиц делают процесс выполнения практической работы увлекательным, позволяют обучающимся использовать свою креативность для решения задач визуализации данных, потому что лица имеют множество параметров, которые можно использовать для описания данных, и, соответственно, предоставляют выбор, какие параметры лиц будут отвечать за те или иные параметры данных.

В практических работах также много внимания уделяется программным средствам для реализации. Обучающимся дается необходимая и достаточная информация для освоения python библиотеки визуализации Matplotlib [3]. С точки зрения освоения библиотеки визуализации лица Чернова являются удачным решением, так как для построения лица в указанной библиотеке недостаточно стандартных функций, упрощающих построение различных графиков – для визуализации лиц необходимо использовать более низкоуровневый программный интерфейс, дающий больше контроля исполнителю, и позволяющий строить различные геометрические фигуры на графике. Архитектура Matplotlib логически разделена на три уровня, которые можно рассматривать в виде стека по уровню абстракции. Каждый уровень, находящийся выше по цепи взаимодействия, является потребителем программного интерфейса звеньев более низкого уровня, при этом звенья ниже уровнем не обладают информацией об уровнях с более высокой абстракцией. Библиотека состоит из трех основных слоев от низкого уровня к высокому – Backend, инкапсулирующий в себе концепт «бумаги», низкоуровневое рисование и обработку пользовательского ввода; Artist, упрощающий взаимодействие с Backend; и pyplot для написания пользовательских скриптов. Библиотека Matplotlib обладает исчерпывающей документацией [4] и примерами использования [5]. Таким образом, имея понимание архитектуры библиотеки Matplotlib, используя ее более продвинутые программные интерфейсы и комбинирование геометрических фигур предоставляемых Artist (в частности Patch) можно построить визуализацию лица, причем с достаточно высокой вариативностью, что делает процесс более интересным для обучающихся.

Так как в работе обучающимся необходимо осваивать более низкоуровневый программный интерфейс, в методических указаниях приводятся объяснения и примеры построения лиц Чернова для данных с меньшим количеством параметров, которые показывают основной алгоритм построения визуализации, оставляя при этом достаточное пространство для принятия собственных решений. Изучение и самостоятельный поиск информации являются одними из ключевых навыков для решения задач в глобальном информационном обществе. В методических указаниях приведены ссылки на документацию по библиотеке Matplotlib, и, если обучающийся пожелает построить визуализацию более продвинутого уровня, для которой необходима дополнительная информация, он сможет это сделать, изучив дополнительные материалы из документации по библиотеке.

Предложенные практические работы были апробированы в 2022/2023 учебном году, и вызвали значимый интерес учащихся к практической части курса «Визуализация данных». Стоит отметить, что в основном обучающиеся ограничиваются изучением методических указаний и лекционного материала, так как он является достаточным для выполнения поставленных задач. Для стимулирования самостоятельного углубления в тему курса студентам предложен практический кейс, в основе которого лежат задания VASTChallenge. Задания, сформулированные в рамках данного конкурса, максимально близко приближены к реальным аналитическим задачам, и решая такой кейс, студенты

имеют возможность полностью погрузиться в процесс анализа данных как методами визуализации, так и машинного обучения.

Освоение и свободная работа с использованием современных информационных технологий, к которым относится и визуализация данных в задачах обработки и интерпретации данных, приобретение навыков поиска оригинальных профессиональных решений, позволяет высшей школе готовить как специалистов, способных решать актуальные научные задачи современного информационного общества, так и стать пользователями ресурсов глобального информационного общества, подготовленными для решения актуальных многофункциональных современных задач.

Список литературы:

1. Herman Chernoff (1973). «The Use of Faces to Represent Points in K-Dimensional Space Graphically». // Journal of the American Statistical Association 68 (342): 361–368.
2. Jan Fazlagić, Windham Loopesko, Leszek Matuszak, Rigby Johnson. Perspectives and good practices in visualization of knowledge about public entities // Information Visualization Techniques in the Social Sciences and Humanities / Grzegorz Osinski, VeslavaOsińska. – IGI Global, 2018. – С. 199–214. – ISBN 9781522549918.
3. Библиотека визуализации Matplotlib. URL: <https://matplotlib.org/> (дата посещения 10.01.2023).
4. Документация библиотеки Matplotlib. URL: <https://matplotlib.org/stable/index.html> (дата посещения 13.01.2023).
5. Примеры использования библиотеки Matplotlib. URL: https://matplotlib.org/stable/plot_types/index.html (дата посещения 14.01.2023).

M. D. Kuznetsov

Approach to the formation of practical tasks for conducting practices in the discipline "Data Visualization" within the framework of the master's program "Information Systems and Technologies"

Saint Petersburg Electrotechnical University, Russia

Abstract. *The paper discusses the experience of forming methodological materials for the discipline "Data Visualization" using the technique of visualizing multidimensional data of Chernov's faces and the Matplotlib software package for students of the master's program in the field of "Information Systems and Technologies". The choice of the used visualization technique and visualization library is substantiated. The strengths and weaknesses of the visualization technique are given, the ways of stimulating the process of studying the discipline are described.*

Key words: practical exercises; guidelines; assignments; master's program; data visualization; Chernoff faces