THE ROLE OF PYTHON PROGRAMMING COURSE IN NATURAL SCIENCES

Sidita Duli
University "Luigj Gurakuqi", Shkoder, Albania, sidita.duli@unishk.edu.al

Abstract: The study of Natural Sciences often leads to Data Science, a trending topic among skilled professionals and organizations. Data Science is the newest field that promises to revolutionize industries, businesses, government, health care, and academia. A data scientist should own mathematical expertise, strong business and technology skills.

With the development of higher education, teaching methods, courses content, the technology used in the learning process, and the curriculum system should reform to adapt to the needs of society. Some data scientists have studied Computer Science. Meanwhile, many others come from Statistics, Mathematics, Physics, or Biology backgrounds and don't have programming skills. This article shows the role of programming in Natural Sciences study programs in Universities and suggests the programming course content.

Many data scientists and software developers use Python to predict outcomes, automate tasks, streamline processes, and offer business intelligence insights. Python provides many functions to Data visualize and data analysis. Python syntax is easy to understand and write, which makes it very popular.

With the advantages of Python, the Natural Sciences students create the main skills in data science, mathematical computations libraries, and data analysis.

The methodology used in this research is the SciPy, the main Python library that includes linear algebra, integration, optimization, and statistics modules. Another library is NumPy. It is a perfect tool for scientific computing and performing basic and advanced array operations. Matplotlib is a standard data science library that generates data visualizations, such as two-dimensional diagrams and graphs (histograms, scatterplots, non-Cartesian coordinates graphs).

This article aims to emphasize the role of the Python programming language in all the study programs in Natural Sciences. It discusses the course content of Python Programming, the main ideas to cover to create the basic knowledge of Data Science. It suggests that Natural Sciences students should master these three libraries: NumPy, SciPy, and matplotlib. It describes each of these libraries with main functionalities.

Keywords: Data Science, Natural Sciences, Python programming

1. INTRODUCTION

With the development of higher education, teaching methods, courses content, the technology used in the learning process, and the curriculum system should reform to adapt to the needs of society. Natural Sciences study programs often lead to Data Science, a trending topic among skilled professionals and organizations. Data Science is the newest field that promises to revolutionize businesses to government, health care, and academia. A data scientist should own mathematical expertise, strong business and technology skills.

Python is a popular object-oriented programming language created by Guido van Roussum in 1990. Later on, many developers and programmers contributed to its growth. Its popularity has been escalated in recent years in many different domains, including Data Science. Many data scientists and software developers use Python to predict outcomes, automate tasks, streamline processes, and offer business intelligence insights.

One of the benefits of using Python is that it provides numerous ready-to-use robust libraries that developers maintain actively. Besides freely available libraries, there are multiple other reasons why Python is such a popular programming language, not only for Data Science but also for Machine Learning, Artificial Intelligence, game development, and Business Applications. The Python hype is likely to grow for many more years to come, and recently, many top U.S. universities have switched to Python for teaching introductory programming classes. [1]

The study [2] shows the exploration and reform on the school-enterprise course Python Data Analysis. It shows the requirement of enterprises and determined the teaching content accordingly.

In the qualitative research [3], the students have demonstrated greater ability and interest in peer collaboration and shared problem-solving. They are far more comfortable learning without having a clear, definable answer to challenges.

The article [5] shows the results of a visit through the websites of the computer science departments. There is a diversity of courses related to Python, as the main computational tool: 'Introduction to Computer Science Programming in Python' in Massachusetts Institute of Technology, U.S.A., 'Python for Data Analysis & Visualization' in Stanford University, U.S.A., 'Python for Developers' in Carnegie Mellon University, U.S.A.,

'Introduction to Computational Thinking with Data' in the University of California Berkeley, U.S.A. and 'Python: Introduction of Absolute Beginners' in University of Cambridge, U.K.

In the related papers [3], [6], the students who participated in the questionaries are not Computer Science students. The study [3] concludes in many programming education opportunities, which lead to the improvement of the computational thinking abilities of the students.

Computational thinking refers to understanding problem-solving, system design, and human behavior based on computer science's basic concepts and principles [7]. It is the basic knowledge that future talents must possess. In particular, for the development of computing algorithms, majors not related to IT must find innovative items in their fields of study, overcome their fear of technology, and develop the ability to communicate with IT professionals. Computational thinking can be developed and improved through computer programming language education.

The programming languages are very diverse, but it is desirable to approach them as a programming language that is interesting for beginners, easy to learn, and capable of writing various applications. Python programming language is one of the appropriate programming languages. Python is used as a pure programming language and as an intermediate language to connect modules written in other languages [8].

Python course is a new programming course at the University of Shkodra, in the Computer Sciences study program. This research aims to relate Python with other study programs. With the advantages of Python, the students of non-computer professionals using Python as the first programming language is advisable. To make the students acquire programming ability, teachers need to use effective teaching methods and establish new teaching modes. [6]

The related studies show the increase of the Python courses, which can also teach other Computer Science courses, such as computer architecture, operating systems, computer networks, databases, algorithms and data structures, machine learning, cloud systems, and more.[9]

2. MATERIALS AND METHODS

The research is based on the library that offers Python, part of the scientific Python ecosystem (SciPy), a collection of Python libraries for mathematics, science, and engineering.

SciPy includes NumPy, SciPy library, and matplotlib. [10]. Other important components of this ecosystem are:
- Pandas, which stands for Python data analysis and provides versatile data structures
- SymPy, which stands for symbolic Python, and is very useful for symbolic mathematics and algebra.
- scikit-image, which includes functions for image processing.
- scikit-learn, which contains tasks for machine learning.

The two interactive IDEs that we suggest to work with SciPy are IPython or Jupyter Notebook.[11]

Of course, a professional Data Scientists should master each of the libraries above, but this research focuses only on three of them.

NumPy library

NumPy stands for Numerical Python. It is one of the most useful scientific Python programming libraries and supports large multidimensional array objects and various tools to work with them. NumPy arrays are called ndarray or N-dimensional arrays, and they store elements of the same type and size.

NumPy is very popular for its high performance, efficient storage, and data operations as arrays grow in size.

SciPy library

SciPy library provides utility functions for optimization, stats, and signal processing. Most new Data Science features are available in SciPy rather than Numpy.

matplotlib

The matplotlib is the oldest and most widely-used Python library for data visualization, created by neurobiologist John D. Hunter. It provides a MATLAB-like interface but with the advantage of being free and open source. The researcher can control almost every setting in the matplotlib chart: image size, dpi, line width, color and style, axes, axis and grid properties, text and font properties.

The research aims to use these libraries in the course materials of Python programming and relate them to the concrete course projects in Natural Sciences.

3. RESULTS AND DISCUSSIONS

The Python Programming course in the Natural Sciences study programs should introduce the basic theories and methods for data acquisition, processing, modeling, and analysis by using the above Python libraries. The course is suitable for every student who wants to perform data analysis.

After finishing the Python Programming course, students will gain the basic data analysis and data visualization ability and computational thinking applied in all scientific fields in which they are engaged.

Generally, a full-cycle data science project includes the following stages:
- Data Gathering & Exploring;
- Data Analysis & Modeling;
- Communication & Deployment.

Exploratory data analysis (EDA)

Most of the data science projects start with Data Gathering and Exploring. Thus, the students should master exploratory data analysis skills. In this course topic, the student will gain basic knowledge in NumPy and matplotlib. By using NumPy, the student builds summary statistics for the dataset. Meanwhile, by using matplotlib, the student creates histograms, scatter plots, and other dataset visualizations, which are important tools for understanding the dataset and identifying possible outliers.

The topic of EDA project should be related to the study program. It might analyze weather data, environmental data, and so on.

Data visualizations

Data Visualizations, such as histograms and scatter plots, are part of the data exploration and analysis steps. The library used in this topic is mainly matplotlib.

The topic of the project should focus on visualization and advanced plots. A project example is building a Covid-19 Dashboard. A more advanced project might be to visualize climate change data with Python.

Tabular data analysis

In business and science, lots of data are organized in tables. Thus, one of the most important data science topics should include working with tabular data using Python.

There are many popular datasets with tabular data. Depending on the study program, the project might be predicting how the bounce of a basketball is affected by Air Pressure, analyzing the Water Temperature at various depths, etc.

Sentiment analysis

The most frequent text data analyses are the reviews. There is a special interest that the public health community uses these sources for opinion mining. The sentiment analysis might have in focus patient feedback classification.

The project of this topic requires some advanced machine learning tools, but Python makes sentiment analysis pretty straightforward. Some project examples include Sentiment Analysis on Food Reviews, Life Expectancy Prediction, Breast Cancer Analysis.

Anomaly detection

Anomaly Detection is the next important topic of Python. The libraries used in this topic are matplotlib and NumPy. The project of this topic might focus on the classification of astronomical objects – which, despite being named like a classification problem, actually turns out to be another anomaly detection problem.

In Table 1, there is a completed list of project ideas of what Natural Science students can implement during the Python Programming course.

Table 1: List of projects in Python applied in Natural Sciences

Topic	Projects in Python
Exploratory data analysis	Weather Data Analysis
Exploratory data analysis	Environmental Data Analysis
Data visualizations	Building a Covid-19 Chart
Data visualizations	Visualize Climate Changes Data
Tabular data analysis	Basket Ball bounce by Air Pressure
Tabular data analysis	Water Temperature by the depth
Sentiment analysis	Food Reviews
Sentiment analysis	Life Expectancy Prediction
Sentiment analysis	Breast Cancer Analysis
Anomaly detection	Classification of astronomical objects

The project list contains actual problems in Data Science, such as the Covid-19 Chart or Climate changes. The libraries SciPy, NumPy, and matplotlib are an important part of the solution of the projects mentioned in this section.

4. CONCLUSIONS

The introduction of Python programming in the study programs of Natural Sciences is an optimization of the actual curricula and a requirement. Nowadays, we cannot perceive a Scientist without the abilities of Data Gathering, Data Analyzing, and Data Visualizing. Such skills are a demand in Biology, Chemistry, Physics, and Nursery.

This study lists a set of projects, which correlate with Data Science and Natural Science studies. They include basic concepts in Python programming, implementing the main libraries SciPy, NumPy, and matplotlib.

REFERENCES

Arif, T.M. (2020). Introduction to Deep Learning for Engineers: Using Python and Google Cloud Platform.

Liu, X., & Xu, H. (2019). School-Enterprise Cooperation on Python Data Analysis Teaching. IEEE Xplore. https://doi.org/10.1109/ICCSE.2019.8845524

Lee, Y., & Cho, J. (2017). The Influence of Python Programming Education for Raising Computational Thinking. International Journal of U- and E-Service, Science and Technology, 10(8), 59–72. https://doi.org/10.14257/ijunesst.2017.10.8.06

Fagan, B. J., & Payne, B. (2017). Learning to Program in Python – by Teaching It! Proceedings of the Interdisciplinary STEM Teaching and Learning Conference, 1(1). https://doi.org/10.20429/stem.2017.010109

Sotomayor-Beltran, C., Segura, G. W. Z., & Roman-Gonzalez, A. (2018). Why should Python be a compulsory introductory programming course in Lima (Peru) universities? 2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA). https://doi.org/10.1109/ica-acca.2018.8609808

Kui, X., Liu, W., Xia, J., & Du, H. (2017). Research on the improvement of Python language programming course teaching methods based on visualization. IEEE Xplore. https://doi.org/10.1109/ICCSE.2017.8085571

Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33. https://doi.org/10.1145/1118178.1118215

Brennan, K., & Resnick, M. (2021). New frameworks for studying and assessing the development of computational thinking", In Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada, pp. 1-25.

Kantaria, M., Basilaia, G., Dgebuadze, M., & Chokhonelidze, G. (2020). Applying a new teaching methodology to university programming language courses.

Lee, W.-M. (2019). Python machine learning. Wiley.

Pajankar, A. (2021). Practical Python Data Visualization: A Fast Track Approach To Learning Data Visualization With Python.