TOWARDS A BRIGHT FUTURE WITH DATA SCIENCE

Jamshid Ibadullayev

Author: A student of Karshi State University
jamshidibadullayev91@gmail.com

Annotation: *What is Data Science? Data Science is the acquisition, preparation, analysis, visualization and delivery of data. It is a multifaceted field that includes scientific analysis, software algorithms, and complex processes. In this article, we will provide information about what knowledge and technologies those who are interested in the field of Data Science, who want to go in this direction in the future, can learn.*

Keywords: *Artificial intelligence, Data Science, framework, mathematical model, Programming language, Python, Visualization*

INTRODUCTION

In recent years, interest in Data Science, i.e. Data Engineering, which is the basis of modern Artificial Intelligence, has been growing. In this article, we will provide information about what knowledge and technologies those who are interested in the field of Data Science, who want to go in this direction in the future, can learn.

The need for experts working with data has increased to such an extent that even future businesses are being talked about depending on representatives of this field. What is Data Science? We would answer this question as follows. Data Science is the process of obtaining data from information sources using scientific methods and algorithms, and applying it to business. It's no secret that any business and organization make their decisions based on the available data, and the field that does and studies this is Data Science.

Today, Data Science is used in almost all data industries - retail, e-commerce, finance, healthcare, education, and telecommunications. In addition, Google, Apple, Facebook, Amazon and several other large companies that have great power in the Internet world are doing great things using Data Science. These companies turn to Data Science to develop their business and implement their decisions through accurate and accurate data.

Remember that Data Science means "science" in its name, that is, a branch of science. Modern artificial intelligence programs are based on analyzing gigabytes and even terabytes of information, extracting useful information from them, and finally directly creating mathematical models.

Mathematics and statistics: While modern frameworks are very helpful in rendering mathematical formulas programmatically, you need to have an idea of how the formula works and what results to expect from it. So the knowledge required in this step is:
- Functions, equations, formulas and graphs;

• Statistics: average value, median value, variance, correlation, etc.;

• Probability theory: probability calculation, Bayes theorem, conditional probability, different distributions (uniform, normal);

• Linear Algebra: vectors, matrices and operations on them. Special types of matrices;

• Discrete mathematics and data structures: sets, stacks, queues, graphs, arrays, hash tables and trees;

• Fundamentals of mathematical analysis: derivative, integral.

Remember, what is required of you is not to be a mathematician, but to have an idea of where the above things are used and how they work.

Programming language: Of course, as an expert in the field, it is required to master one of the programming languages suitable for Data Science. Today you have two choices:

• R: Used for direct data manipulation, various mathematical and statistical calculations and analysis. If you're strong in math and want to get straight into data analysis, R is your programming language of choice.

• Python: A high-level, general-purpose programming language. Python is used to create everything from simple programs to complex calculations and systems.

If your goal is to go through mathematical calculations and create programs based on artificial intelligence, you cannot find a better language than Python. Python is a language that is easy to learn, easy to program, and most importantly, rich in ready-made functions, libraries, and frameworks for Data Science.

Today, companies like Google, Facebook, Tesla, IBM, Amazon are also using Python to create their artificial intelligence programs.

Data warehouse: In order to work with data, we need to know how to save data somewhere, refer to it when necessary, filter and extract the part we need from the data.

Today, using SQL (and similar) data warehouses is the most popular way to store data. So you also need to have minimal knowledge of working with SQL databases. Including accessing the repository, retrieving data from it based on a certain parameter or filter, sorting data, uploading them back to the repository, etc.

Data Collection and Processing: Data engineering is a huge field and one must learn first-hand data collection and processing before entering the field. Data can be in the form of text, tables, images, audio/video, archival documents, etc. Your primary task will be to align the data, fill in gaps, delete erroneous data, etc. before processing it.

To do this, you need to know at least a few programs that work with different types of data. For example, Excel for working with tables, programs that convert pdf or image documents into text, etc.

In order to process the data directly in Python, it is advisable to familiarize yourself with the pandas library. Also, an important part of this step is to find or gather

the necessary data sets for your project by consulting different sources. Such collections are usually called datasets, and there are many sites online that collect datasets on various topics.

Visualize data: It's important to present data in the form of beautiful graphs, rather than dry tables and numbers, to make it easier to understand and present to clients or colleagues. It also makes data analysis much easier.

Machine learning: Now we come to the point of interest. Now that your data is ready, you can train your computer and create various useful programs with this data. Machine Learning is a very large area and consists of such large fields as supervised learning (education under management), unsupervised learning (independent education), reinforcement learning.

Your first step starts with building various prediction systems based on supervised learning (dollar prices, house prices, weather, etc.). Then you will step into unsupervised learning and learn to work on problems such as classification and segmentation. So the libraries and frameworks you should learn in this step are: numpy, scikit-learn, keras, scipy, tensorflow or pytorch.

Natural Language Processing (NLP): Direction of working with natural languages. This step is completely optional and is intended for those who want to take their level up another notch. This direction is aimed at creating intelligent programs that understand human language by analyzing speech and texts. "Ok Google" or "Siri" services on modern phones and computers, smart chatbots on various sites belong to this category of programs.

This route can be very easy (English, Russian) or very complicated (Uzbek) depending on the language you are programming in.

The demand for Data Scientists will continue to grow. In 2012, the demand for data workers increased by 650%, and by 2026, this indicator is expected to appear in the job market of 11.5 million jobs (U.S. Bureau of Labor Statistics).

REFERENCES:

1.   Data Science yo'l xaritasi - Texnoman.uz

2.   https://towardsdatascience.com/a-new-plot-theme-for-matplotlib-gadfly-2cffc745ff84

3.   Natural Language Processing with PyTorch: https://www.amazon.com/_/dp/1491978236?tag=oreilly20-20

4.   Sun'iy Intellekt |Python(Pytorch): https://mansurcompai.github.io/AI_with_PythonPytorch_Uz/

5.   Machine Learning by Andrew Ng: https://www.coursera.org/learn/machine-learning

6.   Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (pdf file, 31mb)

7.   Pandas kutubxonasi (ingliz tilida): https://pandas.pydata.org/

8.   Pandas kutubxonasi boʻyicha qoʻllanma (ingliz tilida): https://www.w3schools.com/python/pandas/default.asp

9.   Farkhod Dadajonovdan SQL darslari: https://youtube.com/playlist?list=PL1538BF8E4A3E5D56

10.   Maʼlumotlar tuzilmasi va algoritmlar onlayn kursi: https://mohirdev.uz/courses/algoritmlar/