

CREATING A DATA SCIENCE ROADMAP AND ANALYSIS**Tojimamatov I.N***Teacher Fergana State University***Olimov A.F***Student Fergana State University***Khaydarova O.T***Student Fergana State University***Tojiboyev M.M***Student Fergana State University*

In recent years, interest in Data Science, which is the basis of modern Artificial Intelligence, has been growing. In this article, we provide a detailed roadmap for those interested in Data Science. We will consider what knowledge and technologies those who want to go in this direction in the future should learn. Whenever possible, we also recommend a manual, tutorial or online course course for each step.

MATHEMATICS AND STATISTICS

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.

In general, 70-80% of the time to create artificial intelligence is data analysis and the remaining 20-30% is direct programming. Therefore, a person entering this field is required to have minimal knowledge of mathematics, statistics and probability theory.

While modern frameworks help a lot 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, dispersion, 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.

Time: How long it takes to master the above knowledge is up to you. My personal recommendation is to look at the lessons on that specific topic when you are faced with a specific math problem, rather than spending a lot of time learning them.

SOURCES:

- Mathematics school course video lessons: <https://www.youtube.com/c/>
- Khan Academy lessons in Uzbek: <https://uz.khanacademy.org/>
- "PROBABILITY THEORY AND MATHEMATICAL STATISTICS" book (pdf download).
- Data structure and algorithms online course: <https://mohirdev.uz/courses/algoritmlar/>

1. PROGRAMMING LANGUAGE

Of course, as a specialist 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 beyond mathematical calculations and create AI-powered programs, you won't 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.

Time: Do not spend more than 1-3 months.

Source:

- Fundamentals of programming in Python: <https://python.sariq.dev>

3. DATA WAREHOUSE

In order to work with data, we need to know how to save data somewhere, refer to it at the right time, 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.

Time: Don't spend too much time on this step, it shouldn't take more than 2-3 weeks to master the minimum knowledge mentioned above.

- Sources: Farkhod Dadajonovdan SQL darslari: <https://youtube.com/playlist?list=PL1538BF8E4A3E5D56>

- SQLga oid qo'llanmalar: <https://www.tutorials.uz/tutorial/mysql>

4. VERSION CONTROL (Git)

One of the must-know skills for any developer is version control. As you work on a Data Science project, your code is constantly updated. Both the data used during the project and the functions used to process it may improve over time or, on the contrary, encounter "bugs". In such cases, special version control systems will make your work a thousand times easier to save every step, every change of your project and return to an older version when necessary, or to branch the project and go in a new direction.

In addition, with the help of such systems, you can permanently save your project in the cloud and refer to it from anywhere at any time, or share it with colleagues and clients.

The most popular version control system today is Git and its GitHub page.

- Duration: 1 week long

- Sources: Farkhod Dadajonovning video

darslari: https://youtube.com/playlist?list=PL_WK6W0Gn1I5GclCgBTsq4pW2rKV9cK-F

- Saidbek Arislonov. Git va Github:

<https://www.youtube.com/watch?v=JtVnOZ26XHA>

5. DATA COLLECTION AND PROCESSING

Data engineering is a huge field and it is imperative to learn how to collect and process data directly 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.

And for processing data directly in Python, I recommend getting to know 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.

Recommendation: If you are new to Data Science, start by learning how to work with ready-made datasets. For this, see the pandas library mentioned above.

Time: Approximately 2–3 weeks to familiarize yourself with the pandas library.

Manbalar:

- Pandas library (in English): <https://pandas.pydata.org/>

- Pandas Library Guide (in English):

<https://www.w3schools.com/python/pandas/default.asp>

- Kaggle datasets repository: <https://www.kaggle.com/datasets>

- Google Dataset search engine: <https://datasetsearch.research.google.com/>

6. DATA VISUALIZATION

It is very important to present the data in the form of beautiful graphs, not in the form of dry tables and numbers, in order to make the data understandable and present to clients or colleagues. It also makes data analysis much easier.

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

The most convenient and perfect libraries for creating graphs today are Matplotlib and seaborn libraries.

Recommendation: It is enough to choose one of the above libraries. Matplotlib is feature-rich, and seaborn is easy to draw eye-pleasing graphics.

Time: 2-3 weeks

Sources:Manbalar:

- Matplotlib qo'llanma (ingliz tilida): https://www.w3schools.com/python/matplotlib_intro.asp

- seaborn qo'llanma: <https://www.tutorialspoint.com/seaborn/index.htm>

7. 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.

Source: <https://towardsdatascience.com/machine-learning-types-2-c1291d4f04b1>

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.

8. DEEP LEARNING

Deep Learning is actually a part of the above Machine Learning direction and aims to create models that find solutions to complex problems using Neural Network. In particular, programs such as human face recognition, classification of objects in images, voice recognition, working with natural languages (chatbots) are created with the help of Deep Learning.

- This step also uses the aforementioned Creas, Tensorflow, and PyTorch frameworks.

- Time: 3-4 months

- Sources:Deep Learning Specialization:

<https://www.coursera.org/specializations/deep-learning>

- Deep Learning with PyTorch (pdf kitob, 45mb)

- Deep Learning with Python: <https://www.manning.com/books/deep-learning-with-python>

9. 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.

REFERENCES:

1. ДэвиСилен, АрноМейсман, МохамедАли. Основы. Data. Science и BigData. И. Python и наука о данных. ПИТЕР Санкт-Петербург Москва Екатеринбург Воронеж. 2017.

2. Mamasidiqova, I., Husanova, O., Madaminova, A., & Tojimatov, I. (2023). DATA MINING TEXNALOGIYALARI METODLARI VA BOSQICHLARI HAMDA DATA SCIENCE JARAYONLAR. Центральноеазиатский журнал образования и инноваций, 2(3 Part 2), 18-21.

3. Tojimatov, I.N., Mamalatipov, O.M., & Karimova, N.A. (2022). SUN'IY NEYRON TARMOQLARINI O'QITISH USULLARI Oriental renaissance: Innovative, educational, natural and social sciences, 2(12), 191-203.

4. Nurmamatovich, T. I. (2021). RAQAMLI IQTISODIYOTNING GLOBALLASHUV JARAYONIDA IQTISOD TARMOQLARIDA QO'LLANILISHINING ASOSIY YO'NALISHLARI. H34 Наука и инновации в XXI веке: Материалы Международной, 291.

5. Tuuchievich, B. M., & Nurmamatovich, T. I. (2021). ЖАМИЯТДА РАҚАМЛИ ИҚТИСОДИЁТ. H34 Наука и инновации в XXI веке: Материалы Международной, 189.

6. Kizi, A. Z. I., & Nurmamatovich, T. I. (2021). ZAMONAVIY DASTURLASH FANINI O'QITISHDA PYTHON DASTURLASH VOSITALARI YORDAMIDA AMALIY DASTURLAR YARATISHNING AHAMIYATI. H34 Наука и инновации в XXI веке: Материалы Международной, 264.

7. Tojimatov, I. N., Mamalatipov, O. M., & Karimova, N. A. (2022). SUN'IY NEYRON TARMOQLARINI O'QITISH USULLARI.

8. Abdulaxadov, N., Saminjonov, S., & Tojimatov, I. (2023). MA'LUMOTLAR VA AXBOROTLARNI VIZUALIZATSIYA QILISH USULLARI, INTERAKTIV MEKANIZMLAR. Евразийский журнал технологий и инноваций, 1(4), 7-18.

9. Tojimatov, I., Mirkomil, M. M., & Saidmurod, S. (2023). BIG DATANING TURLI SOHALARDA QO'LLANILISHI. ОБРАЗОВАНИЕ НАУКА И ИННОВАЦИОННЫЕ ИДЕИ В МИРЕ, 18(6), 61-65.

10. Tojimamatov, I., & Doniyorbek, A. (2023). KATTA HAJMLI MA'LUMOTLAR AFZALLIKLARI VA KAMCHILIKLARI. ОБРАЗОВАНИЕ НАУКА И ИННОВАЦИОННЫЕ ИДЕИ В МИРЕ, 18(6), 66-70.
11. Tojimamatov, I., & Xurshidbek, R. (2023). KATTANAJMLIMALUMOTLARNIQAYTAISHLASHDAQOLLANILAYOTGAN TEXNOLOGIYALAR: NOSQL, MAPREDUCE, HADOOP, ERP, SAPNOSQLTEKNOLOGIYASI. ОБРАЗОВАНИЕ НАУКА И ИННОВАЦИОННЫЕ ИДЕИ В МИРЕ, 18(6), 54-60.
12. Abdusalomovna, T.D. (2023). TEXT MINING. European Journal of Interdisciplinary Research and Development, 13, 284-289.
13. Ne'matjonov, F.F., Jahongirova, J.J., Murodov, B.S., & Tojimamatov, I.N. (2023). CREATE DATA CUBE WITH MS EXCEL. European International Journal of Multidisciplinary Research and Management Studies, 3(03), 77-86.
14. Kimyonazarova, D., Ne'matjonova, D., Ergasheva, B., & Tojimamatov, I. (2023, March). KATTA MA'LUMOTLAR BILAN ISHLASHDA HADOOP ARXITEKTURASI. In Международная конференция академических наук (Vol. 2, No. 3, pp. 96-99).
15. T Akhmedov, SM Otajonov, Y Usmonov, MM Khalilov, N Yunusov // Optical properties of polycrystalline films of lead telluride with distributed stochiometry//Journal of Physics: Conference Series 1889 (2), 022052; 41; 2021
16. Akhmedov, SM Otazhonov, MM Khalilov, N Yunusov, U Mamadzhanov, // Effective dielectric permeability and electrical conductivity of polycrystalline PbTe films with disturbed stochiometry//Journal of Physics: Conference Series 2131 (5), 052008; 29; 2021
17. SM Otazhonov, N Yunusov, B Qakhkhorova // Deformation characteristics of PbTe-Te Polycrystalline films // Science and world, 103; 22; 2022
18. SM Отажонов, Н Юнусов//Қаххорова Б//ДЕФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ПОЛИКРИСТАЛЛИЧЕСКИХ ПЛЕНОК PbTe-Te//Деформационный наука и мир 3; 19; 2022
19. SM Отажонов, МХ Рахмонкулов, ПИ Мовлонов, Н Юнусов// Влияние термообработки на фотоэлектрические свойства гетероструктуры Cu₂-xTe-CdTe//Science 89; 19; 2021
20. SM Otazhonov, KA Botirov, MM Khalilov, N Yunusov // IN PHOTONSENSITIVE THIN FILMS CdTe: Ag AND PbTe //Science and World International scientific journal 6 (94), 11-16; 18;
21. SM Отажонов, КА Ботиров, НЭ Алимов // Влияние деформации на миграции дефектов в фоточувствительных полупроводниковых тонких пленках // Finland International Scientific Journal of Education, Social Science ...; 12; 2022
22. SM Otazhonov, MM Khalilov, N Yunusov, T Akhmedov, U Mamajanov // DEVELOPMENT OF A TECHNOLOGY FOR OBTAINING POLYCRYSTALLINE

PbTe FILMS WITH IMPAIRED STOICHIOMETRY//SCIENCE AND WORLD, 14. 2013

23. СМ Отажонов, КА Ботиров, ММ Халилов, Н Юнусов, М Абдумаликова // Спектры поглощения поликристаллических пленок РвТе с избытком теллура// EDITORCOORDINATOR, 954; 12; 2021

24. S Otazhonov, N Alimov, P Movlonov, K Botirov, N Yunusov // Photosensitivity control of CdTe-SiO₂-Si-Al heterostructure with deep impurity levels under external factors//Danish scientific journal 1 (37), 35; 12; 2020

25. МК Онаркулов, СМ Отажонов, КА Ботиров, Н Юнусов // Устройство для изучения тензочувствительности в фоточувствительных полупроводниковых пленках//...Universum: технические науки, 55-58; 11; 2020

26. СМ Отажонов, АМ Худойбердиев, БҚ Абдуллаевич, ММ Халилов // Тензочувствительности полупроводниковых пленок с мелких и глубоких примесей при температуре жидким гелием // Universum: технические науки, 28-32; 8; 2019

27. SM Otazhonov, RM Kh, MM Khalilov, KA Botirov, N Yunusov // Effect of group VII elements on strain sensitivity of polycrystalline films PbTe, PbS // European science review, 35-38; 7; 2021

28. Р.Максудов, ШШ Шухратов, ОА Мирзаев, Н Юнусов // Изучения изменений коэффициента жесткости упругой оболочки прядильной установки // VII International Scientific and Practical Conference "SCIENTIFIC HORIZON IN ...; 5; 2021

29. СМ Отажонов, ММ Халилов, Н Юнусов, Т Ахмедов, У Мамаджанов // РАЗРАБОТКА ТЕХНОЛОГИИ ПОЛУЧЕНИЯ ПОЛИКРИСТАЛЛИЧЕСКИХ ПЛЕНОК РвТе С НАРУШЕННОЙ СТЕХИОМЕТРИЕЙ //Наука и мир, 8-14; 5; 2021

30. С Отажонов, М Халилов, Р Бойбобоев, Н Юнусов, У Мамаджанов // ВЛИЯНИЕ ХЛОРА НА ТЕНЗОСВОЙСТВА ТОНКИХ ПЛЕНОК СУЛЬФИДА СВИНЦА// InterConf, 329-333; 2; 2021

31. С Отажонов, К Ботиров, П Мовлонов, Н Юнусов ИЗМЕНЕНИЕ ФОТОЧУВСТВИТЕЛЬНОСТИ ГЕТЕРОСТРУКТУРЫ НА ОСНОВЕ Cu₂-X Те- CdTe ПРИ ТЕРМИЧЕСКОЙ ОБРОБОТКЕ // InterConf; 2; 2021

32. С Отажонов, К Ботиров, Б Раззоков, Н Юнусов // ИССЛЕДОВАНИЕ ЭЛЕКТРИЧЕСКИХ И ФОТОЭЛЕКТРИЧЕСКИХ СВОЙСТВ ФОТОПРЕОБРАЗОВАТЕЛЕЙ НА ОСНОВЕ ГЕТЕРОПЕРЕХОДА Cu_{2-x}Te–CdTe // InterConf; 2; 2020

33. SM Otazhonov, M Onarkulov, K Botirov, N Yunusov, U Mamadzhanov //Device for studying tenze sensitivity in photosensitive semiconductor films // Universum technical science 2 (71); 2; 2020

34.С Отажонов, Н Алимов, П Мовлонов, К Ботиров, Н Юнусов // УПРАВЛЕНИЕ ФОТОЧУВСТВИТЕЛЬНОСТИ ГЕТЕРОСТРУКТУРЫ CdTe-SiO₂-

Si-Al С ГЛУБОКИМИ ПРИМЕСНЫМИ УРОВНЯМИ ПОД ДЕЙСТВИЕМ ВНЕШНИХ ФАКТОРОВ // DanishScientificJournal, 35-37; ;

35. СМ Отажонов, НЭ Алимов, КА Ботиров, Б Раззоков, М Рахмонкулов // ОЗДАНИЕ ПРЕОБРАЗОВАТЕЛЯ ИЗЛУЧЕНИЯ НА ОСНОВЕ ПЛЁНОЧНОЙ ГЕТЕРОСТРУКТУРЫ p-CdTe-ZnSe // Наука и мир, 11-14

36. ММ Ахмедов, АМ Худойбердиев, Н Юнусов, Н Маматова // О факторах повышения уровня знаний студентов на основе требований инновационной образовательной среды//Universum: психология и образование, 4-6

37. С Отажонов, Р Эргашев, Н Юнусов // ИЗУЧЕНИЕ ПОВЕРХНОСТНОЙ РЕКОМБИНАЦИИ ГЕТЕРОПЕРЕХОДОВ НА ОСНОВЕ pCdTe-nCdS и pCdTe-nCdSe//SCIENCEANDWORLD 26

38. СМ Отажонов, ПИ Мовлонов, Н Юнусов, УМ Мамажонов // РАЗРАБОТКА ТЕХНОЛОГИИ ПОЛУЧЕНИЯ НИЗКООМНЫХ БАЗОВЫХ СЛОЕВ A2B6 МЕТОДОМ ТЕРМОВАКУУМНОЙ КОНДЕНСАЦИИ И ХИМИЧЕСКИМ ОСАЖДЕНИЕМ // Universum: технические науки, 5-9

39. С Отажонов, К Ботиров, Н Юнусов//ВЛИЯНИЕ ДЕФОРМАЦИИ НА МИКРОПОТЕНЦИАЛЬНЫХ БАРЬЕРОВ В ПОЛИКРИСТАЛЛИЧЕСКИХ ПЛЕНКАХ CDTE ЛЕГИРОВАННЫХ СЕРЕБРОМ // InterConf

40. СМ Отажонов, НЭ Алимов, КА Ботиров, Н Юнусов, ША Ибрагимова// БЫСТРОДЕЙСТВУЮЩИЙ ФОТОДЕТЕКТОР НА ОСНОВЕ ГЕТЕРОСТРУКТУРЫ P-CDTE-SIO2-SI С ГЛУБОКИМИ ПРИМЕСНЫМИ УРОВНЯМИ// EDITORCOORDINATOR, 624

41. YN Otazhonov S.M., Rakhmonkulov M.Kh., Movlonov P.I.//EFFECT OF HEAT TREATMENT ON THE PHOTOVOLTAIC PROPERTIES OF THE SI 2-TE-CDTE HETEROSTRUCTURE // Science and World International scientific journal 1 (№89), 22-27

42. С Отажонов, Н Юнусов, К Ботиров, Б Раззаков // СОЗДАНИЯ ФОТОПРЕОБРАЗОВАТЕЛЕЙ НА ОСНОВЕ ГЕТЕРОПРЕХОДА CU2 ТЕ-CDTE И ИССЛЕДОВАНИЕ ИХ ФОТОЭЛЕКТРИЧЕСКИХ СВОЙСТВ // ЗАМОНАВИЙ МИКРОЭЛЕКТРОНИКАНИНГ РИВОЖЛАНИШИДА ФАН,ТАЪЛИМ ВА ИННОВАЦИЯ.

43. С Отажонов, Н Юнусов // ЭЛЕКТРИЧЕСКИЕ СВОЙСТВА ТЕНЗОЧУВСТВИТЕЛЬНЫХ ПЛЕНОК PbS ПОД ДЕСТВИЕМ ЛАЗЕРНОГО ОБЛУЧЕНИЯ // ОПТИЧЕСКИМ И ФОТОЭЛЕКТРИЧЕСКИМ ЯВЛЕНИЯМ В ПОЛУПРОВОДНИКОВЫХ МИКРО-И.

44. С Отажонов, Н Юнусов, К Ботиров, Н Алимов // ЭЛЕКТРОННЫЙ ПЕРЕНОС НОСИТЕЛЕЙ ЗАРЯДА В ГЕТЕРОСТРУКТУРЕ CDTE-SIO2-SI-AL С ГЛУБОКИМИ ПРИМЕСНЫМИ УРОВНЯМИ ПОД ДЕЙСТВИЕМ ВНЕШНИХ ФАКТОРОВ // ОПТИЧЕСКИМ И ФОТОЭЛЕКТРИЧЕСКИМ ЯВЛЕНИЯМ В ПОЛУПРОВОДНИКОВЫХ МИКРО-И

45. С Отажонов, Н Юнусов, К Ботиров, Б Раззаков // CREATION OF A RADIATION CONVERTER BASED ON FILM HETEROSTRUCTURE P-CDTE-ZNSE // Science and world 87 (№11), 11-15

46. СМ Отажонов, Н Юнусов, Б Каххорова // ДЕФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ПОЛИКРИСТАЛЛИЧЕСКИХ ПЛЕНОК PbTe-Te // SCIENCEANDWORLD, 27