Group of
Horribly
Optimistic
STatisticians

Intro

Intro to Data Science

Maksymilian Norkiewicz & Jędrzej Ogrodowski

Jędrzej Ogrodowski
- 3rd year CS student
- R&D Intern, Demant
- Interested in applied ML, LLMs, analytics

Maksymilian Norkiewicz
- 3rd year AI student

Communication and attendance

For communication:

Discord

Attendance:

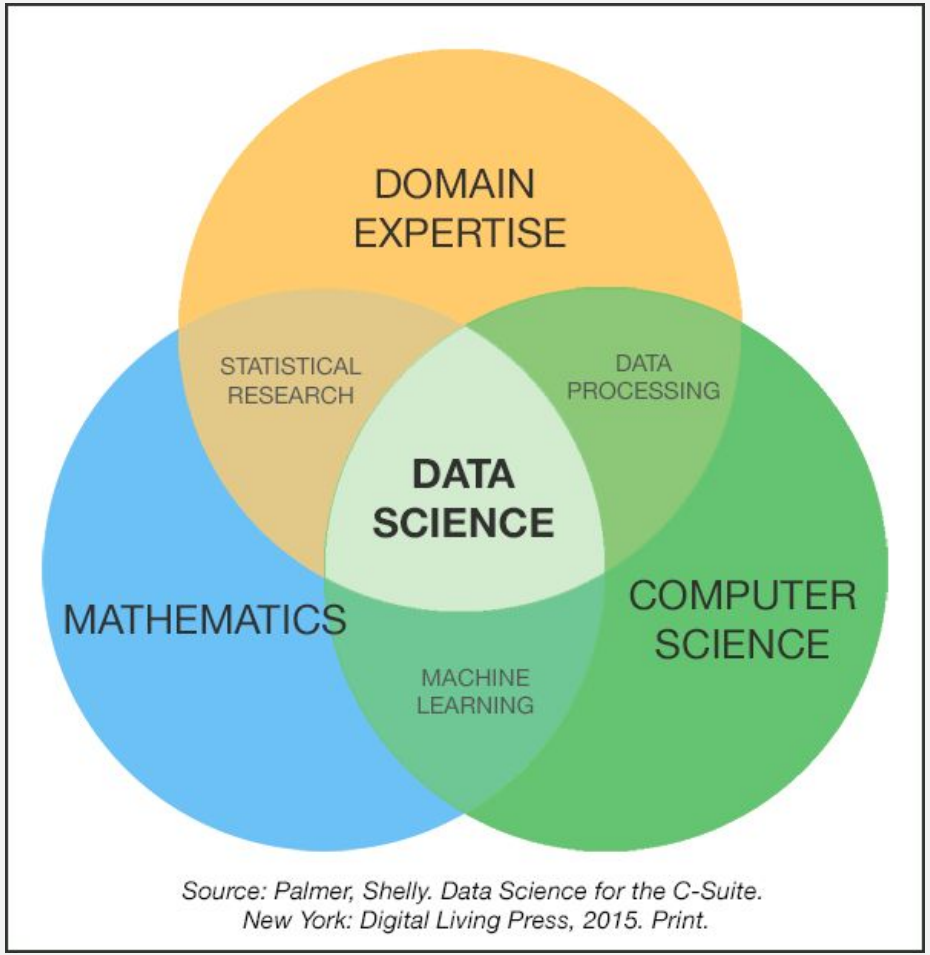
You can record your attendance in an Excel file which you can find on Discord

So, what is Data Science?

What is Data Science

Data Science is scientific field that uses statistic, data processing, machine learning algorithms and visualizations to extract knowledge from data and solve business problems.

Our goal is to bring order to chaotic data, present it in meaningful way, develop predictive models and draw insightful conclusions.



Source: Palmer, Shelly. Data Science for the C-Suite. New York: Digital Living Press, 2015. Print.

Current Data Science roles

Obowiązkowe

Machine Learning | Data science | AI | Python | Cloud | GenAI | LLM | Angielski (C1)

Mile widziane

TensorFlow | Kubernetes | MLflow | CI/CD | AWS | PyTorch | Docker | RAG | LangChain | LlamaIndex | OpenAI | Huggingface | Azure | MLOps | Computer vision | Polski

TECH STACK

English B2	Machine Lear... Advanced	Azure Advanced	Python Advanced	GenAI Nice To Have
AWS Nice To Have	LLM Nice To Have	PyTorch Nice To Have	rag Nice To Have	TensorFlow Nice To Have
GCP Nice To Have				

Obowiązkowe

Data science | Machine learning | Deep learning | Python | scikit-learn | SQL | PyTorch | MLOps | CI/CD | Communication skills | Angielski (B2)

Let’s create a plan!

How to solve typical task

1. Define problem
2. Gather data
3. Clean data
4. Analyse data
5. Visualize and draw conclusions
6. Develop and deploy models

Section plan

1. Intro to “Intro to Data Science” (28.10)
2. Data manipulation (04.11)
3. Statistics pt. 1 (18.11)
4. Data visualization pt. 1 (25.11)
5. Statistics pt. 2 (02.12)
6. Data visualization pt. 2 (09.12)
7. Exploratory data analysis (16.12) + project subject
8. Prediction models (13.01)
9. **Project presentation (20.01)**

After Intro

After this section, you will have a solid foundation in data collection, transformation, and exploration, as well as a good intuition for creating meaningful stories with your data.

Most importantly, you'll be ready to start your own projects and build on these fundamentals in other GHOST sections.

Handouts

All handouts, presentations, exercises and projects are available on our sections Github.

To access it you will need to enter link at Discord, **star repository** (this is very important step), fork it.

Literature

- W. McKinney, “Python for Data Analysis”
- Claus O. Wilke, “Fundamentals of Data Visualizations”
- C. N. Knaflic, “Storytelling with data”
- G. James et al., “An Introduction to Statistical Learning”
- S. Raschka, “Machine Learning with PyTorch and Scikit-learn”
- J. T. VanderPlas, “Python Data Science Handbook”

Thank you