

# Doporučené zdroje: ML/DL



Zde najdeš výběr knih, kurzů a online zdrojů, které lidem z našeho ML a research týmu opravdu pomohly. Neber ho jako checklist nebo seznam znalostí, který bys měl nutně ovládat. Spíš nabízí vhled do toho, jak o machine learningu a deep learningu přemýšlíme a kde se může rozumně zorientovat někdo, kdo se chystá na pohovor do našeho ML/DL týmu.

V Qminers používáme ML a DL na složitá, zašuměná finanční data, kde je signál jemný a předpoklady hrají velkou roli. Proto si ceníme pevných základů v pravděpodobnosti, statistice a modelování, stejně jako dobrého pochopení toho, jak moderní metody strojového učení a deep learningu fungují v praxi. Doporučené materiály tenhle balanc mezi teorií a praktickou zkušeností odrážejí.

Pokud tě machine learning zajímá i mimo standardní benchmarky a baví tě modely, které musí obstát v reálných podmínkách, měly by pro tebe tyhle zdroje být relevantní.

## Co doporučuje Ondřej?



Ondřej Skácel  
Quantitative Analyst

### Software Engineering

- 1** **Fluent Python Book (anebo vhodná alternativa)**
- II **Data Structures**
  2. An Array of Sequences
  3. Dictionaries and Sets
- III **Functions as Objects**
  5. First-Class Functions
  7. Function Decorators and Closures
- IV **Object-Oriented Idioms**
  8. Object References, Mutability, and Recycling
  9. A Pythonic Object
  10. Sequence Hacking, Hashing, and Slicing

## V Control Flow

- 14. Iterables, Iterators, and Generators
- 15. Context Managers and else Blocks
- 16. Coroutines\*
- 18. Concurrency with asyncio\*

## 2 Základy NumPy, Matplotlib, Pandas\*, PyTorch a Pytest.

## 3 Základy teoretické informatiky (předpokládáme, že je znáš z univerzitního studia)

- Algoritmická složitost (big-O notace, amortizovaná vs. průměrná vs. nejhorší složitost)
- Běžné algoritmy a datové struktury
- Několik dní řešení úloh na [LeetCode](#) nebo podobných algoritmických soutěžích

## 4 Základy počítačových architektur\*

## 5 Základy Unix/Linux\*

Materiály označené hvězdičkou\* jsou povinné pouze pro pozici ML engineer.

# Quant Research Brainteasers

Zvykni si na řešení problémů.

<https://brainstellar.com/>

<https://openquant.co/questions>

<https://www.janestreet.com/puzzles/archive/index.html>

# Machine Learning

## 1 Probability, Statistics and Linear Algebra

- Práce s pravděpodobnostními rozděleními (CDF, PDF, E, Var, sčítání a násobení)
- Podmíněná pravděpodobnost, úplná pravděpodobnost
- Základy statistiky, statistická významnost, intervaly spolehlivosti, centrální limitní věta
- Např. [Statistical Inference Book](#)
- Matice, vlastní čísla a vlastní vektory
- Derivace a integrace

## 2 Elements of Statistical Learning Book

- 3. Linear Methods for Regression
- 10. Boosting and Additive Trees
- 16. Ensemble Learning

- 3 Případně vybrané části z [Understanding Machine Learning Book](#)
- 4 **Modelling basics** (z přednášky profesora Yasera Abu-Mostafy na Caltech – “Machine Learning Course”)

- [Bias-Variance Tradeoff](#)
- [Overfitting](#)

5 [Neural Network Basics](#)

6 **Deep Learning Background and Architectures**

Např. <https://www.bishopbook.com/>  
(volitelně <https://www.deeplearningbook.org/>)

- Gradient, běžné loss funkce (MSE, Cross Entropy, ...)
- Trénovací algoritmy (SGD, Momentum, RMSProp, Adam)
- Regularizační a normalizační techniky (model, data, trénovací algoritmus)
- Convolutional nets, LSTM, VAE, GAN, SSM
- [Transformers](#)

7 [Google Research Deep Neural Network Tuning Handbook](#)

8 [Karpathy's Recipe for Training Neural Networks](#)