HOW TO GET

WORK IN DATA SCIENCE

KSENIA GASNIKOVA

WHO AM I

AND WHY I'M DOING THIS

2014-2017 PhD student in ATLAS (DESY, Germany)
2017-2018 Data Scientist at NIVEA, Germany
2019 Research Scientist at INRIA (Group of DL/ML)

Do I have enough expertise to do this talk?

- Had in total 2 years of job search in Germany and France
- Did 21x 45-min interviews with Google within 1 month (and didn't get the job)
- Discussed with ex-collegues their experience
- Read lots of guidelines about getting a job in Data Science

“

I DON’T SHINE IF YOU DON’T SHINE

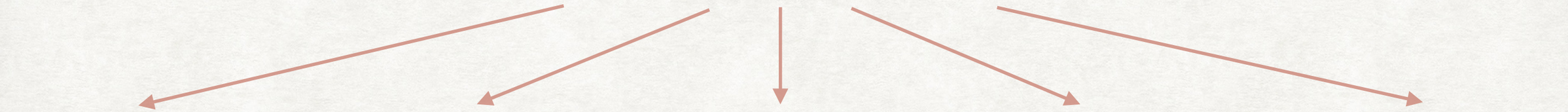
AMINATOU SOW & ANN FRIEDMAN

”

WHAT IS DATA SCIENCE?

"Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it…"

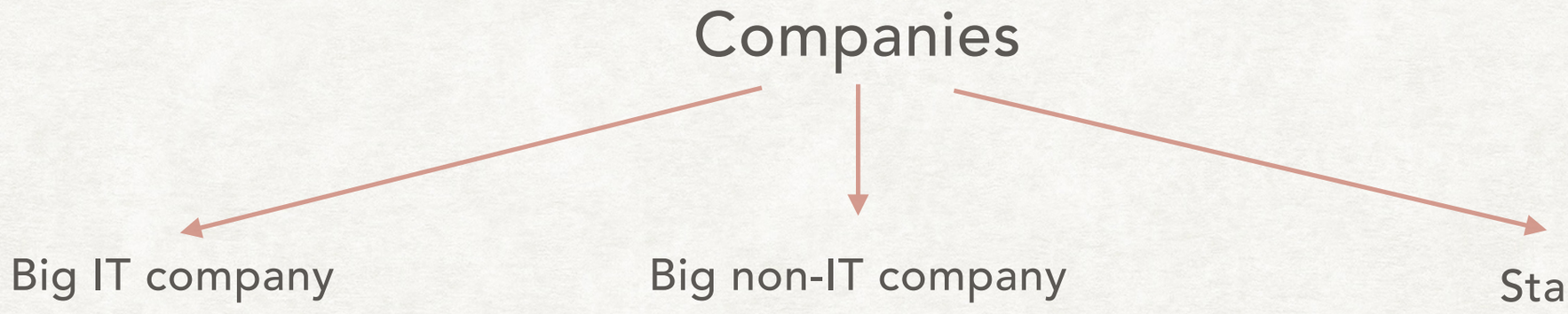
Data Scientists



Product manager	Data Engineer	Data Analyst	Data Scientist	Research Scientist
• Expert in business and consumer needs • Knows a little bit of everything • Spends time in meetings • Uses Excel	• Expert in programming • Knows ML • Spends time optimising ML code, building pipelines • Uses Hadoop, Spark, SQL/ noSQL etc	• Expert in stats and visualisation • Knows ML • Spends time answering questions, building visualisation and reports • Uses Tableau, Spark, R, SaS, SQL, Excel	• Expert in ML/DL/ Stats • Knows programming and visualisation • Spends time trying to find something interesting in Data • Uses Python, R, Keras etc	• Expert in math behind ML/DL • Knows everything • Spends time building state-of-the-art models • Uses Python, R, C++, PyTorch, Tensorflow

WHO WANTS DATA SCIENTISTS?

"Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it…"



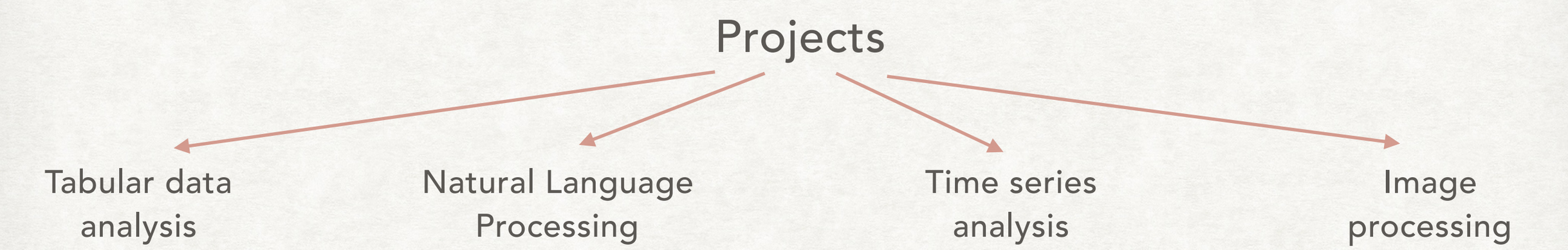
- Interview level nightmare
- Assigned task
- Flexibility
- Lots of effort to teach you
- Looks extremely good in CV

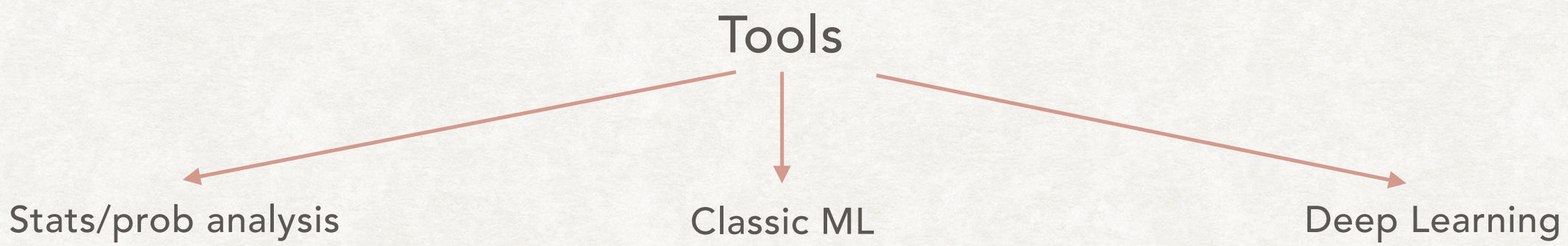
- Interview level easy to middle
- Find your own (unexpected) task
- More flexibility than you want
- Want you to teach them
- Looks good in CV

- Any level of interviews
- Lots of different assigned tasks
- No flexibility
- You will learn on your own, but extremely fast
- Nothing fancy in CV

WHAT WILL YOU DO

It’s usually better to be clear about your areas of interest





HOW TO GET AN INTERVIEW

MY PERSONAL SCALE

PROBABILITY OF GETTING THE JOB

- Kaggle competitions/ML conferences/MeetUp

  If they see that you are amazing - things will go the way easier.
  Also not all of the positions are posted anywhere online

- Friend/colleague referral

  Or they see that somebody in the company knows that you are amazing

- Via recruiter (usually LinkedIn)

  Recruiters will help your CV not to be lost.
  Spend time and make a really good LinkedIn page - it's really beneficial

- Via job posting

  Big companies are getting sometimes up to hundreds of CVs per job posting . It's almost like to get a permanent position in academia

MAKING A GOOD CV

LONG BORING RUSSIAN STYLE SLIDE

Making a really good CV is hard. The more time you spent doing it, the more chances you have to get the job.

Usual problems that I see in (ex-)scientists CVs are:

- Using the same CV for academia and industry
- Listing ROOT and all other e.g. HEP tools that nobody knows about
- Being too precise about their favourite search/cross-section/asymmetry analysis
  —> **NOBODY WILL BE INTERESTED IN THIS !!!**
- Forgetting that you've been a key figure in your analysis and not highlighting that
- Listing all of the talks/presentations that you had since your high school

* Also, don't make your CV in a sophisticated (TeX) template. Many companies pre-process CVs, so all of this fancy formatting is lost and your CV will end up looking like a mess.

HOW TO MAKE A GOOD CV

ADVISES FROM BOOKS

- General
  - Change your CV every time when you are applying to a job
  - Do not use more than 2 pages
  - Check grammar (&spelling!) with a native speaker
  - List your achievements/awards
  - Add outreach activities
- Experience
  - Make your results measurable, eg improved classification by 20%
  - Do not add up too many projects
  - Try to explain this as down-to-earth as you can
- Skills
  - Write only the things that you are comfortable with
  - Do not make the list too big

TYPES OF INTERVIEWS

THAT I HAD

HARDNESS OF INTERVIEW

1. Personality
2. Job fitting
3. Ask your own question
4. Tackling a problem (creativity)
5. Which kind of algo would you use (skills)
6. Derive equation of this ML algorithm
7. Stats/prob/combinatorics
8. Algorithms/programming language

- HR/recruiters will usually ask questions 1-3 (but google has a little bit of 8).
- Phone interviews will be with 1-5 (sometimes 6)
- On-site can have everything, but usually they tell you in advance
- You can also find company-specific questions on Glassdoor

PERSONALITY INTERVIEW

- Think about the most interesting projects that you did
- Fill the grid and remember it by heart

	Project 1	Project 2	Project 3
Challenges			
Mistakes			
Enjoyed			
Leadership			
Conflicts			
Do differently			

- Know your weaknesses

Don’t say that you are perfectionist or working too hard
Say how you are trying to improve

- Just read “Cracking the coding interviews” by Gayle Laakman McDowell since she has a great chapter about this type of interviews

HOW TO DO THIS?

This is a really good example of

- Why you decided to be particle physics
- What have you been doing at your previous work

JOB FITTING

Try to prepare for each interview. I personally try to read through the original job posting and try to understand the main things:

- What skills are they looking for and what can I change in my presentation of my experience to fit their needs?
- That kind of my personal interests can be useful for them?

  I got my job in NIVEA IoT department partially because I had a wearable that they didn't know about

- Which kind of unique perspective can they get out of my expertise as a (particle) physicist?

ASK YOUR OWN QUESTION

Since Data Science “is like teenager sex”, lots of the job postings are not matching the real work that you are going to do there.
It’s your chance to investigate **if this job is good for you**

Examples of possible questions:

- How many people are in the team, what is their experience?
- Which kind of tasks will you be doing?
- What should you achieve?
- Are there any opportunities for promotion?
- How long is this project? What is going to happen after this?
- And many other

PROBLEM/ALGO QUESTIONS

Usually requires not a deep knowledge, but rather something broad. Easier to prepare if you will read a little bit every day. What I like:

- Kaggle blog + newsletter
- Linear Digressions podcast
- Data Science weekly newsletter
- Towards Data Science blog
- Google Data Science blog

ML EQUATIONS

It’s better to remember by heart derivation of basic algorithms like:

- Ordinary Least Squares

It can be complicated, trust me.
Try to say 5 assumptions of linear regression
and which assumption can be easily excluded

- Support Vector Machines
- Naive Bayes
- Your favourite clustering algorithm
- Gradient descend (or its extensions)
- Boosting algorithm

STATISTICS/PROBABILITY

Do not assume that particle physicists know A LOT about statistics. Read through some of the books and try to remember things like:

- Basic test statistic like t-test, f-test, ANOVA, parametric/non parametric tests, bootstrap
- Remember definition of confidence interval and p-value by heart since it's a good GOTCHA!
- Practice basic things like coin flip and permutation

You can find lots of this type of questions in this blog

ALGORITHMS/PROGRAMMING LANGUAGE

It can be a really rare type of questions, but many ML-based companies take this type of interviews to crazy levels.

Generally:

1. Remember basic sorting algorithms, their differences, time and space complexities
2. Practice algorithmic questions, always remember to debug the solution manually after finishing
3. Try to make mock interviews before the real interview

Learn:

1. “Cracking the coding interviews” Gayle Laakman McDowel
2. Interviewbit for trial interviews
3. Algorithms by Stanford University

HOW TO LEARN ML

THINGS THAT I LIKE

- Machine Learning by Andrew Ng - must
- Deep Learning by Andrew Ng - good intro to DL
- FastAI courses about DL - good courses for coders
- Advanced Machine Learning - lots of math
- Intro to statistics - good and fully free course
- School about bayesian methods in ML/DL
- Google cloud specialisation, or this

NEED MORE EDUCATION?

There are also many different free schools for PhDs who want to change to Data Science/Deep Learning research. Usually after this programs employment is guaranteed

What I personally know about:

1. Insight Data Science

   You will need an US work permission

2. Reply Data Incubator

   In Germany and Italy, work permission in country needed
   Program do not cover so much Deep Learning

3. Lots of AI residency programs by big companies

   Mostly for people who want to go into Deep Learning research

OFFER?

Don’t have so much experience with this part of job searching process, so only basic advices

1. Before saying how much you want - look at Glassdoor, Linkedin statistics about salary or ask people working in company

2. Negotiate salary, even if it’s hard!

   A friend of mine knew that he can easily negotiate salary X euros/year
   So he asked for X euros/year, got X - 5 000 euros/year
   Don’t do it like this!

3. Don’t forget to ask for salary increase after probation period =]

ADDITIONAL QUESTIONS?

If you have additional questions or just want to share you success story, don't hesitate to reach me by:

- email : ksugasnikova@gmail.com
- linkedin: https://www.linkedin.com/in/ksenia-gasnikova/

GOOD LUCK WITH YOUR JOB SEARCH!