The Data Science Impostor

DR MATEUSZ JURCZYŃSKI

MATEUSZ.M.JURCZYNSKI@GMAIL.COM

Who am I?

➢Pure maths PhD from Lancaster

➢Learned bits of Python & machine learning as a hobby

➢Joined the CGI Technical Graduate Scheme in September 2016

➢Joined Innovative Technologies Ltd as Data Scientist in July 2017

➢Uses data to smash friends in board games

CGI

Innovative Technology
INTELLIGENCE IN VALIDATION

Data Science – what is it?

Josh Wills
@josh_wills

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

9:55 am - 3 May 2012

Data Analysis & Diagnostics
(Microsoft Excel, mean & variance, SQL)

Statistical modelling
R

Machine learning
Python, R, Matlab, C++

Big Data
Hadoop, Spark, MapReduce

Cloud computing, deep learning, computer vision, data modelling, databases, software engineering, software development, version control, data visualisation…

Machine learning?

➢ Modelling + simulation using a computer

➢ Curve fitting can be phrased as a machine learning problem

➢ Supervised and unsupervised

➢ Well-known "easy" algorithms:
  ➢ Support Vector Machines
  ➢ Decision Trees & Random Forests
  ➢ Naïve Bayesian

➢ ARIMA!

Why does it have to be snakes (Python)?

- Python or R?
  - R has more „actual stats"
  - Python is more of a „general programming language"
  - Opinions on which one has more machine learning differ...

- Cool stuff in Python
  - Package for everything
  - ENORMOUS amount of resources
  - Is „hyped"
  - Allows for deployment

- Your comfort is most important

- You need a prototyping language

Resources

➢ Data Science & Machine learning
  ➢ https://www.kaggle.com/ - Data Science problems & competitions
  ➢ https://www.kdnuggets.com/ - „10 most popular algorithms for X" and the like
  ➢ https://machinelearningmastery.com/ - General ML blog with many useful examples

➢Programming
  ➢ https://www.codingame.com/ - solve game-like programming problems...
  ➢ https://www.hackerrank.com/ - solve programming problems (markets itself for interviews)
  ➢ https://www.codecademy.com/ - learn programming language from scratch
  ➢ https://projecteuler.net/ - solve maths problems via programming
  ➢ https://repl.it/ - in-browser IDE

➢ General
  ➢ https://news.ycombinator.com/ - imagine „Technology Reddit" with less noise.
  ➢ https://www.edx.org/, https://www.coursera.org/, https://eu.udacity.com/ - multiple courses on X in Data Science, Y in Statistical Modelling, Machine Learning…

Getting the job in the first place

➢ DS recruitment agencies
  ◦ Pitfalls

➢ Grad schemes

➢ Data science-esque directions
  ◦ Programming in general
  ◦ Data Analysis
  ◦ Finance
  ◦ Research

LinkedIn profile

➢IT IS IMPORTANT

➢Smart photo

➢Get connections

➢Data title for what you want to be

➢Location for where you want to be

➢RESPOND TO MESSAGES!

➢``Open to new opportunities" button

Job titles   Current or Past ▾

Data Scientist +

+ Data Analyst, + Head Of Science, + Head,

Locations   Current   ▾

Manchester, United Kingdom +

+ London, United Kingdom (2,832),

Skills

+ Skills and expertise or boolean

+ R (61), + Data Analysis (87),

Companies

+ Companies or boolean

+ First Derivatives, + UCAS, + IBM,

Year of Graduation

+ Add graduation year range

Schools

+ Schools attended

+ Imperial College London (1), + UCL (1),

Industries

+ Candidate industries

+ Information Technology and Services (30),

Keywords

+ Profile keywords or boolean

Employment type

CV

Bullet points rather than long paragraphs

In XYZ company I've been responsible for filing papers and making coffee. I used people's history of coffee consumption to inform my coffee delivery habits.

XYZ company – Jan 2010 - June 2017
- Filing papers
- Coffee making
Personal qualities
- Data-driven

PROOFREAD

Don't cram into 1 page, if it needs to be 2 (or 3) it's fine!
- Although if it starts exceeding this much maybe you should be more concise...
- Universities seem to teach the 1 page format? - **NO**

Replicate buzzwords from job advert

DATES

Choose what's relevant!

Interview questions

➢ Be professional, come prepared – notebook & pen never hurt!

➢ Be prepared for the "usual questions": why you're interested in the job, the area, what techniques have you used already (e.g. programming language), tell of a situation where you had to display skill X (leadership, resolve a conflict, etc.)
  o STAR pattern: Situation, Task, Action, Result

➢ Have a question for the interviewer!

➢ Briefly describe your (PhD / degree) research.

➢ What is the structure of a neural network?

➢ Give an example of an activation function.

➢ What is the difference between generative and descriptive models?

➢ What is the algorithmic complexity of… (quick sort, bisection search, etc.)

➢ Starting from a 3D cube corner and moving along each edge with equal probability, what is the expected number of steps to go from one corner to the one "across"?

➢ Given a set of numbers from 1 to n with one number missing, what algorithm would you use to find the missing number?

➢ What is object oriented programming?

➢ What is the expected number of rolls of a fair 6-sided die to have every result appear at least once?

Companies / areas to look at

➢ Consultancy companies
  ➢ CGI
  ➢ Deloitte
  ➢ CapGemini

➢ Anything with a strong „social media" vibe
  ➢ LADBible

➢ Quantitative finance
  ➢ Gresearch

➢ Gaming companies, mobile and otherwise

➢ Startups

Thank you very much for your kind attention.

Questions?

Mateusz.m.jurczynski@gmail.com