Agenda

• Who is Constantine Stokolos?
• My Career Path
• Data science : Consulting vs Industry
• How to improve? Resources
• QnA
Who is Constantine Stokolos?

- Senior – Machine Learning Engineer - Advisory DnA - EY
- Ex IBMer
- MBA, Georgia Southern
- Master of Data Science (Applied Statistics), Kennesaw State
- [linkedin.com/in/constantinestokolos/](https://linkedin.com/in/constantinestokolos/)

**Industries:**
- CPG
- Retail & Fashion Goods
- Utilities
- Fraud Prevention
- Postal
- Pharma

**Sectors:**
- Public
- Government
- Commercial

**Clients:**
- Nike Inc
- Southern Company OpCos
- USPS
- Social Security Administration
- Pfizer
## Career path

<table>
<thead>
<tr>
<th>Year</th>
<th>PROGNOS INC</th>
<th>SOUTHERN COMPANY</th>
<th>IBM</th>
<th>EY</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>SAS</td>
<td>SQL</td>
<td>Python</td>
<td>R</td>
<td>SQL</td>
</tr>
<tr>
<td>2013</td>
<td>SAS</td>
<td>SQL</td>
<td>Python</td>
<td>R</td>
<td>SQL</td>
</tr>
<tr>
<td>2016</td>
<td>Python</td>
<td>Spark</td>
<td>Spark</td>
<td>Python</td>
<td>Spark</td>
</tr>
<tr>
<td>2018</td>
<td>R</td>
<td>SQL</td>
<td>SAS</td>
<td>TensorFlow</td>
<td>TensorFlow</td>
</tr>
<tr>
<td></td>
<td>SQL</td>
<td>Python</td>
<td>MxNET</td>
<td>Keras</td>
<td>Keras</td>
</tr>
</tbody>
</table>

Technical skills:
- SAS
- SQL
- R
- Python
- SPSS
- Spark
- Scala
- SAS
- Tensorflow
- Keras
- MxNET
- Scala
- Spark
- TensorFlow
- Keras
- MxNET
Data Science: Consulting vs Industry

Consulting Pros
- Higher Compensation
- Lower technical requirements
- Relatively easy to get in
- Consider yourself a ‘product’
- Faster Growth
- Exposure
- Various industries
- Traveling

Consulting Cons
- Sales. Sales. Sales!
- More Work. (45-50 hrs +)
- Client comes first
- Lots of miscellaneous activities that take your time
- Poor working conditions
- Traveling
Data Science: Consulting vs Industry

Industry Pros
- More balanced week work hours
- Office space
- Stable environment
- Decisions are not driven by client
- Quality of solutions are usually higher
- More hours on actual development

Industry Cons
- Lower salaries
- Higher technical requirements
- You are an overhead
- Less exposure
- Slower growth
- Same industry, repetitive tasks
Data Science : Consulting vs Industry

Trends

- Interactive Dashboards & Reporting Automation
- Big Data processing : Spark
- Deep Learning
- NLP (Both Deep Learning and Traditional)
- Cloud Computing : AWS, Azure, others
- BI Enterprise solutions
- Fraud and Cyber Security
- Medical : Image Processing, Anomaly Detection, Pattern Recognition, etc
Data Science : Consulting vs Industry

My opinion on how to proceed:

- Start with consulting
- Get familiar with various industries
- Get Exposed to lots of companies
- Switch to industry if consulting is not for you, about Manager and up, or 5+ years of experience
- Stay in consulting if Sales is what you like. Rewards are huge!
- Stay away from time consuming tasks that don’t improve your skill set or boost your resume
Real Life Experience. Must Haves in your Skillset

1. Python:
   1. Core: Numpy, pandas, matplotlib, seaborn
   3. Deep Learning: TensorFlow, Keras, MxNet
2. Spark, SQL:
   1. PySpark
   2. Spark Shell
   3. Spark Mllib
3. R
4. SAS
5. Cloud:
   1. AWS: EC2, EMR, Lambda, S3, others
   2. Microsoft Azure
   3. GPC
6. Hadoop Framework
7. NoSQL & SQL
8. ETL (example Apache Sqoop)

Vitals:
- Linear Algebra
- Stat Methods
- Probability Theory
- Calculus
- Bash (Unix Shell)

Real Life Experience. R, Python, SAS, Others

- Python, SQL & Spark
  - Must have for anything, especially ML, DL
- R
  - when a company tries to switch from SAS or have already some R code in the system
- Scala
  - Native Spark API, master if you want a clean Spark knowledge
- Java
  - Barely used in consulting as development time is too long


R:
- R good for exploratory
- R - powerful statistical modeling
- R doesn’t scale
- R misses some classes and objects
- R No native Deep Learning libraries

SAS:
- Expensive
- Falls behind on availability of new algorithms
- Have their own algorithms engines, black box
- Often you need multiple packages or even software:
  - Eminer
  - Base / EG
  - ETS, HPF, etc
Real Life Experience. Machine Learning vs Big Data

• Ideally you need to be both
• Math background -> Machine Learning Engineers
• Computer Science -> Big Data Architects
• Typically separate positions on a consulting project
• Not so much in the industry
• At the end it is a matter of your preference what you like to do more
Real Life Experience. Continuing Education

- A must do!
- Tools change all the time -> need to keep up and adopt and adjust
- Online Resources:
  - https://www.udemy.com/ Udemy
  - https://www.coursera.org/ Coursera
- Bootcamps:
- Meetups:
- Books:
  - http://www.deeplearningbook.org/ <- FREE
  - http://www-bcf.usc.edu/~gareth/ISL/ <- FREE
- Mock interviews & Resume Critique
  - https://www.evisors.com/
Your turn!