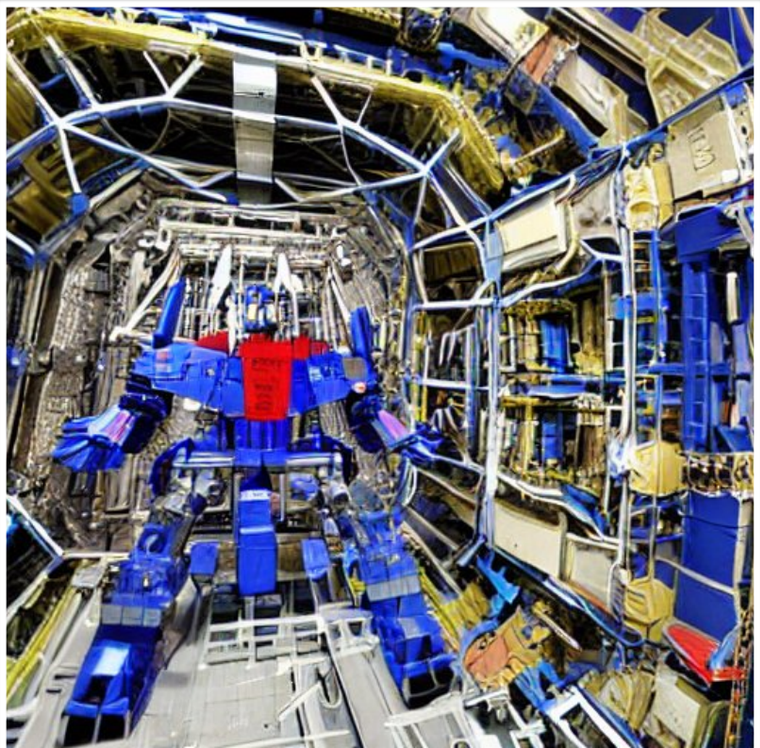
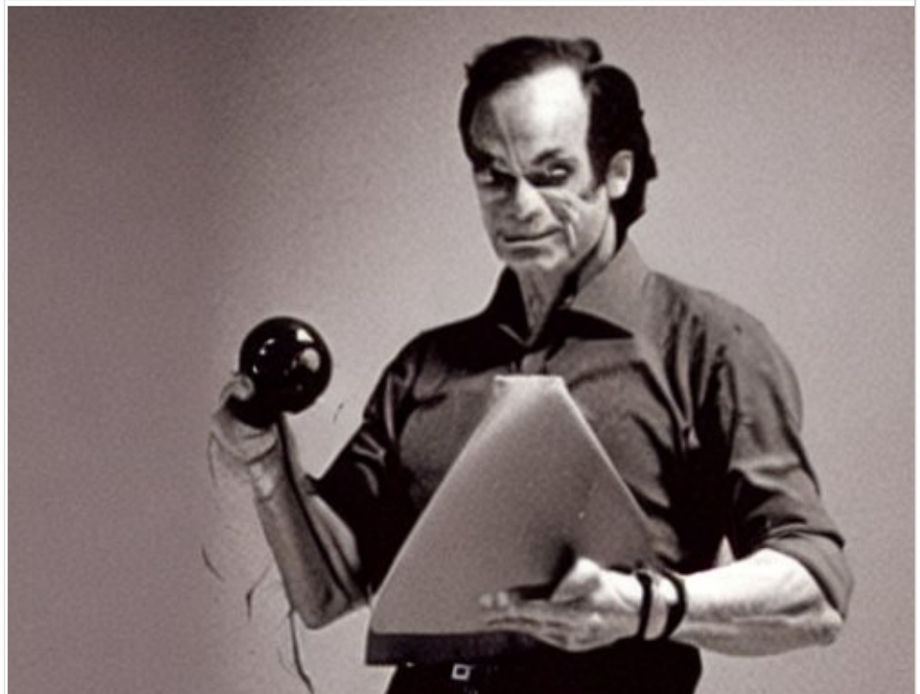
DMPP ANNUAL WORKSHOP

23 Nov 2022

From Particle Physics to Data Science

Hrayr Matevosyan





TRANSFERRABLE SKILLS

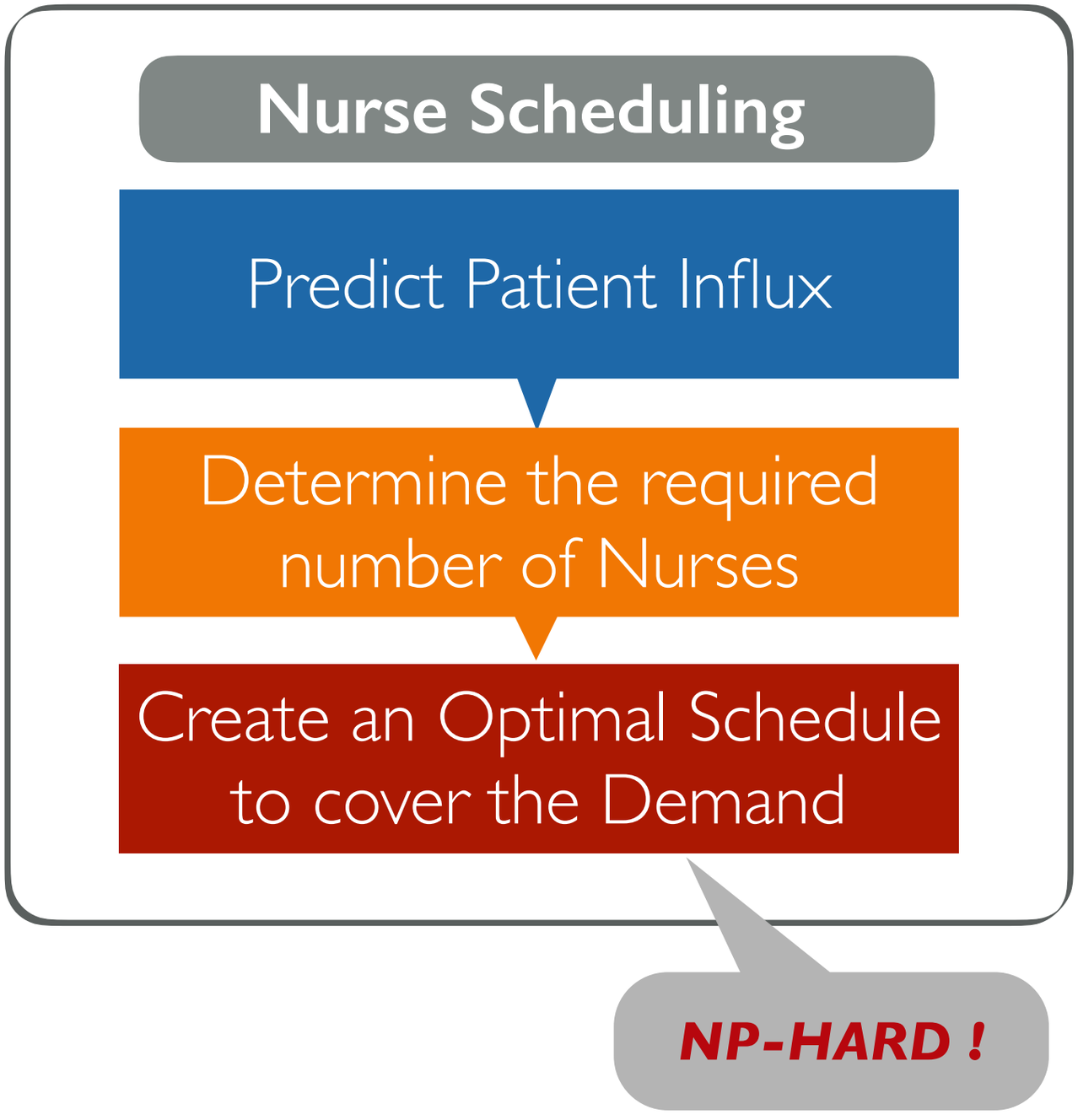
Many pathways in industry and public sector!

- *Research Methods for tackling hard problems.*
- *Strong Maths/Statistics Background.*
- *Computational Skills:* ***Data Analysis, Modelling, HPC, MC Simulations.***
- *Collaboration.*
- *Communication.*

FORECASTING / OPERATIONS RESEARCH

Optimising decision-making problems.

- **Scheduling**
  - *Personnel - airlines, restaurants*
  - *Machines - chemical plants*
  - *Processes (wine/grain blending)*
- **Routing**
  - *Network topology*
  - *Delivery Drivers*
- **Resource Planning**
  - *Cloud compute capacity*
  - *Medical treatments*
  - *Warehouse locations*
- Risk management
  - *Reservoir levels*
  - *National security*



NATURAL LANGUAGE PROCESSING (NLP)

Creating programs that "understand" and process natural language such as text and speech.

- *Examples: Siri, Alexa, Scottish Elevator, GitHub Copilot.*
- *Tasks:*
  - Named Entity Recognition
  - Question Answering
  - Machine Translation
  - Text Generation
  - Sentiment Analysis
  - Summarisation
- ***Extremely Challenging:***

  Often need additional context: "The Silver Arrows were trampled by the Prancing Horse this season".
- ***Recent Breakthroughs:***

  Transformers (2017) and Transfer Learning (2018)

SOLUTION DEVELOPMENT

- *Data Collection / Cleaning / Analysis*
- *Modelling:*
  - Defining ***useful*** metrics
  - Complexity: Quality vs Speed

  - Cost-Efficiency: CPU / GPU / Memory / Runtime
  - Integration and Deployment Complexity: the model should work ***"hands-off"*** as SAAS / On-premise
- *Integration:*

  The model interacts with a complex software system.
- *Production Deployment:*

  Continuous Monitoring / Bug Fixes / Improvements

JOB ADS CAN BE CONFUSING…

- **Role:** *DATA MACHINIST*
- **Skills:**
  - *Have strong expertise in Python, **TensorFlow**, scikit-lear, **Ruby on Rails**, Power BI, **Excel**, baking eclairs, Java Script, fixing motorcycle radiators.*
- **Tasks:**
  - *Create **custom ML models** and host on AWS*
  - *Draw Business insights from Supervised, Unsupervised and **reinforcement learning** using Python*
  - *Design Analytics dashboard with Power BI and create **CI/CD pipeline** with MS flow*
  - *School runs for the CEO.*

“DATA MACHINIST”

Machine Learning

Deep Learning

Vision

NLP

Voice

Data Science

Data Engineering

Analytics / Predictive Analytics

ML Ops

Data Mining

Business Intelligence

Operations Research

Cloud Engineering

CORE SKILLS

- *Programming Languages:*

  Python, Java, Julia, C++, R 🤢

- *Data Processing and Analysis (in Python):*

  Pandas / NumPy / SciPy / Matplotlib / Seaborn / DB(SQL)

- *(Shallow) Machine Learning (Fitting):*

  Scikit-learn / XGBoost / Scikit-optim

- *Deep Learning:*

  PyTorch, TensorFlow, MXNet, Transformers, NLTK, OpenCV

- *Operations Research:*

  Pyomo/Jump/Google OR-Tools/OptaPlanner

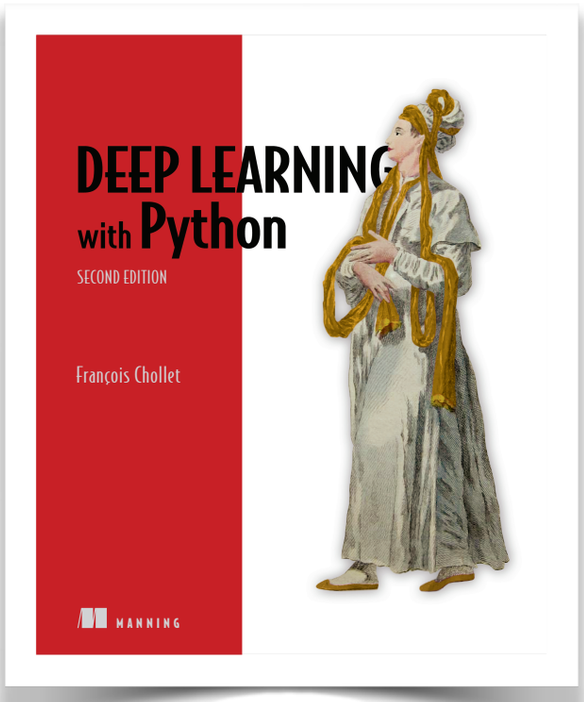
- *Cloud Computing Platforms*

  AWS, Google Cloud, Azure

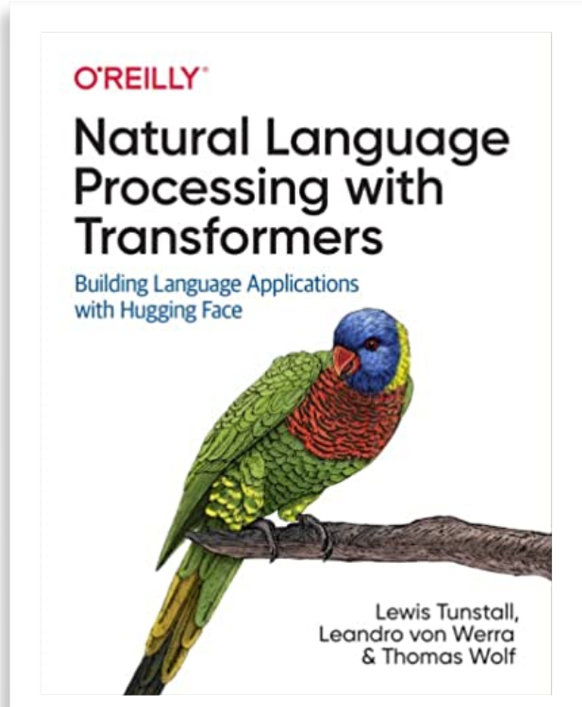
RESOURCES

- *Online Courses:*
  - *Coursera - Many Levels - Machine Learning w/ Andrew Ng*
  - *Youtube - Stanford University Courses*
  - *HuggingFace Courses - https://huggingface.co/course*

- *Books:*
  - F. Chollet



  - L. Tunstall et al.



- *DS Competitions* (Kaggle), *Blogs* (towardsdatascience), *GitHub*

WRITING A CV

Recruiters/HR are the first to read you CV!

Trisha Stamm
E-Mail: trishastamm@gmail.com **Address:** 350 5th Ave New York, NY, 10118
Tel: 401-439-3501

Academic Details

2008 – 2012 - **BS in Statistics, Syracuse University**
- Good at machine learning and data science coursework.
- Completed a senior project to predict waitress tips in a busy Boston restaurant.
- President, AI club.

Programming Languages
- R Studio, Python, PERL
- NoSQL, Hadoop, VB, Cloudera

Work Experience

Date of Joining: 2013-10 to 2017-12
Post: Data Scientist
Organization: Exend Capital Inc.
The scope of work: Fulfilled all data science duties for a high-end capital management firm. Created and presented models for potential holdings to fund managers. Achieved 20% better returns vs historical performance. Created machine learning tools that computed adjusted P/E values.

Date of Joining: 2012-08 to 2013-09
Post: Junior Data Scientist
Organization: Supracy Real Estate Inc.
The scope of work: Collaborated with all team members to optimize Customer Relationship Management database for a high-volume real estate firm. Increased repeat business among real estate investors by 25%. Decreased wasted phone and email time.

Date of Joining: 2012-06 to 2017-12
Post: Freelance Data Scientist
Organization: Various.
The scope of work: NCAAMaster: Beat 200+ statistics professors, students, and professionals in an NCAA pool. FootForest: Used random forest regressor to predict fantasy football scores. Pylearn 2: Regular contributor to Machine Learning project on GitHub.

Hard and Soft Skills
- Machine learning
- Problem solving
- Data visualization
- Modeling
- Critical thinking

Conferences

2016 AAAI Conference on Artificial Intelligence
2014 The Machine Learning Conference

Additional activities
- Article "Machine Learning in the Next 10 Years" cited in Forbes.
- Leader of bi-monthly hiking club.
- Volunteer monthly at the Boylston St. Soup Kitchen.

Trisha Stamm
Data Scientist, Microsoft Certified

Microsoft and Google certified data scientist with 9 years of experience. Seeking to increase data efficiency for Contranix, Inc. Created models to predict stock price with 25% more accuracy than the historical average. Achieved 20% improvement in investment returns.

Experience

2013-10 - 2017-12 **Data Scientist**
Exend Capital Inc.
- Fulfilled all data science duties for a high-end capital management firm.
- Created and presented models for potential holdings to fund managers. Achieved 20% better returns vs historical performance.
- Created machine learning tools that computed adjusted P/E values.
- Predicted stock price 25% better than traditional figures.

2012-08 - 2013-09 **Junior Data Scientist**
Supracy Real Estate Inc.
- Collaborated with all team members to optimize Customer Relationship Management database for a high-volume real estate firm.
- Increased repeat business among real estate investors by 25%.
- Decreased wasted phone and email time by 57%.

2012-06 - 2017-12 **Freelance Data Scientist**
Various
- **NCAAMaster:** Beat 200+ statistics professors, students, and professionals in an NCAA pool. Didn't try to cram basketball into a framework where it didn't fit.
- **FootForest:** Used random forest regressor to predict fantasy football scores.
- **Pylearn 2:** Regular contributor to Machine Learning project on GitHub.

Education

2008 - 2012 **BS in Statistics, Syracuse University**
- Excelled in machine learning and data science coursework.
- Completed a senior project to predict waitress tips in a busy Boston restaurant. Used decision trees, regression models, and SVM.

Microsoft Professional Program for Data Science
- Completed all 10 required courses in Data Science (120 hours).
- Mastered 10 crucial data science skills.

Additional Activities
- Article "Machine Learning in the Next 10 Years" cited in Forbes.
- Volunteer monthly at the Boylston St. Soup Kitchen.

Certifications

Microsoft Professional Program Certificate in Data Science

GCP - Google Certified Professional Data Engineer

Conferences

2016 AAAI Conference on Artificial Intelligence
2014 The Machine Learning Conference

Personal Info

Phone
401-439-3501
E-mail
trishastamm@gmail.com
LinkedIn
linkedin.com/in/trishastamm

Technical Skills

R Studio, Python, PERL, NoSQL, Hadoop, VB, Cloudera — Expert
Machine Learning — Expert
Data Visualization — Expert
Modeling — Expert
Statistics — Expert
Debugging — Advanced

Soft Skills

Problem Solving — Expert
Critical Thinking — Expert
Creative Thinking — Expert
Interpersonal Skills — Advanced

zety

WRITING A CV

Recruiters/HR are the first to read you CV!

- ***Needs to fit the job description!***
- *TOP: Full Contact Info*
  email, phone, LinkedIn, github.
- *Summary of Profile:*
  3-4 sentences of you as a professional - passion/enthusiasm.
- *Experience: Role/dates/ organisation - env/project/ responsibilities/tasks/achievements papers, talks.*
- *Education.*
- *Skills matrix: listing experience.*
- *Publications.*
- *Affiliations, social, hobbies - ok but not essential.*
- *Simple formatting: 1.5-3 pages!*
- *Photo: plain and small.*
- *Key parts of CV reflected in LinkedIn profile.*
- *References when requested.*

Thanks!

Hrayr Matevosyan

hmatevosyan@gmail.com

https://www.linkedin.com/in/hrayr-matevosyan-phd-46a2b512