

Christopher Smith

Data science professional with 8 years of experience empowering organizations with data and insights. My expertise covers multiple areas: defining metrics and extracting insights with imperfect data; general software development including within an object-oriented framework; and the complete machine learning lifecycle from data extraction to deployment.



Professional Experience

2022
I
2019

Senior Machine Learning Analyst

Google

Mountain View, CA

- Built algorithmic protections using machine learning and general programming to remove unwanted content from Google platforms at scale (>10 billion texts, images, videos, and URLs scanned per day)
- Collaborated w/ stakeholders to build, tune, and launch models, including some of the largest product teams within Google.
- Awarded Google Counter Abuse Innovation Award and 10 "Peer Bonuses" given by colleagues
- Spam Text Classifier:** Trained and evaluated ML model end-to-end. Engaged w/ partner team to build dataset. Used NLTK and internal text libraries to clean input text and compute features like part-of-speech. Model achieved 31% improved recall @ 95% precision compared to the prior model.
- Phishing Campaign Prevention:** Key analyst for a dedicated phishing prevention team (e.g., login credential theft). Ran investigations into false-negatives (misses) using SQL+python, addressed phishing attacks on Google products in real time. Implemented anti-obfuscation web-crawls for a large Google team leading to ~160% increase in phishing warnings
- Web Page Abuse ML Classifier:** Identified and pulled data, iterated on features, trained boosted-trees-classifier in Tensorflow, ran live traffic eval (python), and worked with Engineers to deploy the model. ~620K incremental URLs/week flagged for takedown, and ~130K incremental URLs flagged for manual-review representing ~30% increase in coverage.

2019
I
2015

Modeling & Analytics Associate Manager

Accenture Federal Services

Alexandria, VA

- Built machine learning driven solutions to support contracts with multiple large government agencies.
- Key member in two technical demonstrations, contributing significantly to contract wins in excess of \$20 million. Demos required data analysis and high-speed coding with intense time pressure.
- ML Plagiarism Detection:** Co-led team to automate detection of repeated themes/phrases across thousands of documents, supporting fraud investigations within the first week of field use. Applied OOP principles to architect the overall model framework in Python. Developed binary classifier with >90% precision identifying client labeled fraud. Loaded results to ElasticSearch in AWS for client consumption.
- Biographic Entity Resolution:** Designed and built an entity resolution framework in Java for probabilistic matching of biographic records (names, birthdates, document #s, etc.) in search. Applied text similarity techniques including: edit distance, longest common substring, and double metaphone encoding. Launched 3 search relevancy models in Java within microservices leading to 30%, 89%, and 50% reductions in manual workload respectively for critical government verification processes.

2015
I
2014

Quantitative Analyst

Agilex Technologies

Alexandria, VA

- Identification of criminal activity in networks:** Applied concepts from network theory such as shortest-paths, node centrality, and neighborhood detection to identify risky entities in an interaction graph. Analysis was done using network analysis libraries: igraph (R-lang) and JgraphT (Java).

Contact

Asheville, NC

christopher.smith.t@gmail.com

703-402-6457

[Linkedin: in](#)

Education

University of Virginia
B.S. Systems Engineering
2010–2014

Technical Skills

Supervised and unsupervised machine learning; feature engineering, experiment design and analysis

Exploratory data analysis and visualization

General scripting / software development

Strategic Skills

Track record of communicating with multiple stakeholders to execute on complex projects

Extensive presentation experience with technical and non-technical audiences

Experienced in providing technical mentorship to teammates

Toolset

Java, Python, R, SQL

Linux Shell, Javascript, HTML, CSS, C++, golang

Version control (git), distributed file systems

Access

[Online version](#)