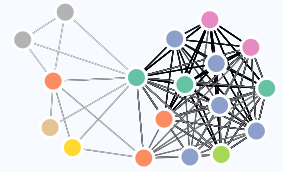


JOSHUA GOLDBERG

Senior applied scientist with expertise in generative AI, abuse prevention, and forecasting. Leads innovative projects at Amazon, combining statistical methods, mathematical modeling, and generative AI to automate complex analyses, uncover strategic insights, and drive impactful operational decisions. Demonstrated experience developing models to proactively detect fraud and accurately forecast global product demand at scale. Proven leader in education and mentorship, regularly instructing/assisting advanced data science and machine learning courses at the University of Chicago and Harvard, and mentoring junior and mid-level scientists.



INDUSTRY EXPERIENCE

Current
|
2020

● Sr. Applied Scientist

Amazon

📍 Seattle, WA

- Lead scientists in generative AI efforts in operations finance. Develop AI strategy and methodology for building and evaluating agentic systems.
- Employed statistical methods, mathematical decomposition techniques, and generative AI to automate analysis of outbound shipping costs, identify financial and operational deviations, and generate actionable insights directly leveraged for operational finance and financial planning. Created generative AI-driven analytical tools with interactive chat capabilities, enabling stakeholders to intuitively deep-dive into model results and financial reports.
- Led the development of machine learning and NLP models to proactively identify and mitigate fraudulent and abusive behaviors from third-party sellers, significantly enhancing the integrity of Amazon's marketplace by preventing phishing, customer diversion, and related fraudulent activities.
- Designed and deployed forecasting models predicting demand for over 500,000 products globally, optimizing inventory management and customer satisfaction.
- Mentored and guided junior and mid-level scientists on statistical techniques, mathematical modeling, time series analysis, software engineering best practices, and career growth.
- Established monthly science reading groups and science office hours, fostering a collaborative environment and ensuring continuous engagement with the latest research and techniques.
- Managed end-to-end lifecycle of science solutions—from concept and model design through implementation and deployment on AWS cloud infrastructure (SageMaker, ECS, Lambda).

CONTACT



josh.goldberg1@outlook.com

 [Github](#)

 [in LinkedIn](#)

LANGUAGE SKILLS

Python

R

SQL

C++

CLI

HTML/CSS

Javascript

2020
|
2017

● **AVP, Lead Data Scientist**

Nuveen

📍 Chicago, IL

- Pioneered end-to-end (execution and experimental design) deep learning time series model for client onboarding; estimated impact of the model was \$1 million net revenue annually that maximized client journey (improvement in client retention, client growth, etc.)
- Built recommendation engine for 150,000 clients in 50+ products
- Presented model/analysis to executive management; results included model adoption by 100+ sales people and a significant increase sales for clients treated by the model
- Conceptualized and created simulation engine that isolated, detected and measured the ROI impact of company sales events

2017
|
2014

● **Senior Equity Research Associate, Financial Services**

Raymond James Financial, Inc.

📍 Chicago, IL

- Built company and industry models using finance and statistical techniques, including regression and discounted cash flows (DCF)



EDUCATION

Current
|
2021

● **STEM Continuing Education**

Harvard, University of Washington

- I actively take STEM courses at different universities to enhance, revisit, or refine my technical skillset. Course topics include computer science, mathematics, machine learning, and statistics.
- Stanford: Convex Optimization 364a with Stephen Boyd (convexity, convex sets and functions, convex optimization problems (linear programs, quadratic programs, semidefinite programs), Lagrangian duality and KKT conditions, optimality conditions, interior-point methods, applications in signal processing, statistics, machine learning); Convex Optimization II 364b (nonsmooth optimization (subgradients, cutting-plane methods), decomposition methods (dual decomposition, ADMM), proximal methods, large-scale and distributed optimization, robust and stochastic optimization, convex relaxations of nonconvex problems, applications to machine learning and control)
- Harvard: Calculus 2 with Series and Differential Equations; Linear Algebra and Differential Equation; Real Analysis; Abstract Linear Algebra; Differential Equations; Systems Programming and Machine Organization (CS61)
- University of Washington: Probability I; Probability II; Linear Optimization; Statistical Inference (STAT 512)
- Edmonds College: CS I, II, III; Courses in C/C++ covering Data structures & algorithms and object-oriented design and programming
- University of Illinois Urbana-Champaign (UIUC): Calculus 1: First course in Calculus and Analytic Geometry

- 2020 ● **M.S. in Applied Data Science**
University of Chicago 📍 Chicago, IL
 • Coursework in statistics, linear algebra, machine learning, and deep learning
- 2013 ● **B.S. in Accounting and Finance**
University of South Florida 📍 Tampa, FL

> SELECTED CODE REPOSITORIES

- 2021 ● **Machine learning decision tree and data frame implementation in C++¹**
Github 📍 Seattle, WA
 • Authored with John Nguyen
- 2020 ● **Generative adversarial network used to generate musical samples²**
University of Chicago 📍 Chicago, IL
 • Capstone project and paper³ authored with Terry Wang and Rima Mittal. Supervised by Yuri Balasanov⁴

👤 TEACHING EXPERIENCE

- Current | 2024 ● **Linear Algebra & Real Analysis**
Harvard Continuing Education, TA 📍 Remote
- Current | 2020 ● **Machine learning, statistics**
University of Chicago, TA 📍 Remote
 • Intro statistics, machine learning, time series analysis
- 2024 | 2022 ● **Python for Data Science**
University of Chicago, Instructor 📍 Remote
 • Topics include introductory and advanced topics in python: variables, logical operators, containers, loops, conditionals, comprehensions, functions, object oriented (basics), advanced data analysis and manipulation with numpy and pandas, model evaluation, parallel computation, and APIs
- 2022 | 2020 ● **MastersTrack Statistics for Machine Learning and Machine Learning Courses**
University of Chicago, Instructor & TA 📍 Remote
 • Statistics course: topics include simple and multiple regression, logistic regression, hypothesis testing, variable transformations. Machine learning course: topics include a survey of machine learning algorithms: kNN, support vector machine, decision tree, random forest, boosted trees, and clustering algorithms

In my free time, I enjoy working with friends, peers, and colleagues on algorithm designs/implementations. One project involved building a data frame and decision tree classes in C++.

I am passionate about teaching and helping others. It brings me joy and satisfication to teach others new skills.

2021
|
2020



Data Understanding via SQL, Databases, and R

University of Chicago, Instructor & TA

 Remote

• Topics include introduction to databases, MySQL, and R