Tanya G. Roosta
1-510-220-0047
Tanya.roosta@gmail.com
US Citizen

SUMMARY

Highly motivated, self-starter individual with more than 12 years of work experience in data science and analysis and machine learning in high tech and quantitative finance industries.

EDUCATION

University of California, Berkeley

2000-2008

Berkeley, CA.

- Ph.D. I Electrical Engineering and Computer Science. Dissertation: *Using Statistical Methods for Anomaly and Attack Detection in Wireless Sensor Networks.* Advisor: Prof. Shankar Sastry

- Masters in Statistics. Thesis: *Convergence Analysis of Reweighted Sum-Product Algorithm.* Advisor: Martin Weignwright

- Masters in Electrical Engineering and Computer Science. Thesis: *Power aware Routing in Wireless Ad-hoc Networks*

- Masters in Financial Engineering, Haas School of Business

- Bachelors in Electrical Engineering and Computer Science

PROFESSIONAL EXPERIENCE

Senior Research Science Manager/Tech Lead, Alexa AI

November 2021-present

Amazon

- Leading a team of scientists and software engineers working on scalable generative AI techniques to increase customer interaction with their Alexa devices through improving their conversational experiences. In this role, I devise strategic roadmap for product release, work with various teams to scope requirements, do research and build the prototype technical solution, and oversee the end-to-end implementation in production.

Senior Research Scientist, Alexa Devices

June 2019-October 2021

Amazon

- Conducting research and experimentations with applying federated learning framework to natural language understanding (NLU) models. The goal of the research is to deploy NLU models on edge devices to increase customer privacy. This work was in collaboration with academia.

- Developing and releasing new features for Alexa mobile devices, such as Echo Buds. The work requires working with various teams to ensure the end-to-end pipeline is setup and runs smoothly. Recent new features include: find the location of an object using a tracker device, and Fitness calories tracker.

- Performance improvement for deep neural network (DNN) models used in the natural language understanding production pipeline. Methods include experimenting with different DNN architectures, and use of semi-supervised learning.

Lead Machine Learning Scientist

March 2018-May 2019

SumUp Analytics (Fintech startup)

- Conducting research and developing the code and REST APIs (offered on AWS API Gateway) for: topic modeling, text summarization, document recommendation, sentiment analysis, and streaming text analytics.

- Managing scientists and engineers working on the full technical stack to ensure a scalable and reliable feature development.

- Working with prospect clients to develop proof of concept for their needs.

Data Science and Machine Learning Lead- Director   May 2016-March 2018

Moody's Analytics

- Leading a variety of engagements for banks and financial institutions, as both quantitative model lead and team manager. The projects covered a spectrum of quantitative risk measurement products and services. Example projects included: credit risk modeling for commercial wholesale portfolio, commercial real estate, residential mortgages, counterparty credit risk.

Associate Director-Quantitative Researcher   May 2014-Aug 2016

Moody's Analytics

- Responsible for improving the country-specific RiskCalcTM Plus model. The model is used to determine the credit risk of private firms based on the financial statements. The modelling process includes: extensive data cleaning, imputation, visualization, univariate analysis of financial ratios for default prediction, multivariate model building and calibration, out of sample testing, cross validation, and predictive power analysis.

Quantitative Risk Analyst   June 2012-April 2014

Federal Reserve Bank of San Francisco-Economic Research Department

- Participating in supervisory exams as the lead quant analyst on bank holding companies regulated by the Federal Reserve Bank of San Francisco including Wells Fargo, Union Bank of California, and Zions Bank.

- Participating in system-wide projects within the Federal Reserve such as the Comprehensive Capital Analysis and Review (CCAR) and BASEL II quantification exams.

- Carrying out empirical research on credit risk modeling, systematic risk indicator using Credit Default Swap (CDS) spreads, and Commercial Real Estate (CRE) modeling and analysis. The research included: data gathering using Bloomberg, RCA, FRED, FR Y-9C and 14A bank schedules, data cleaning, building time series models.

Senior Associate-Financial Engineering and Analytics   Oct 2011-May 2012

PriceWaterHouseCoopers

- Pricing and valuation of complex financial derivatives and contingent claims for the purposes of auditing, strategic assessment and advisory, mergers and acquisitions, and compensation agreements. The asset classes covered included: equity, debt, options, warrants, futures, swaps, convertible securities, embedded derivatives.

Quantitative Researcher-Systematic Equity Portfolio   Oct 2010-Sept 2011

Allianz Global Investor Capitals

- Responsible for the development of statistical risk model for the emerging markets using principal component analysis (PCA)

- Development of a strategy based on the minimum variance portfolio and testing the strategy which led to the creation of AllianzGI U.S. Managed Volatility Fund (NGWAX)

- Developing an expert system for risk management, change point analysis, regime-switching GARCH models, Goodness-of-fit tests for copula, and application of Kalman filter to Fama French factor model

Software Engineer   Aug 2008-Feb 2010

Cisco Systems

- Developed software in C language for the Group Encrypted Transport Virtual Private Network (GETVPN) solution Collaborated on writing the functional specification document developed for the SmartGrid project

Software Engineer (Graduate Intern)   June 2007-Sept 2007

SRI International

- Developed software in C language for the Group Encrypted Transport Virtual Private Network (GETVPN) solution Collaborated on writing the functional specification document developed for the SmartGrid project

---

TEACHING EXPERIENCE

Lecturer at University of California, Berkeley   July 2022-present

Teach Statistics and machine learning in the Master's in Data Science (MIDS) program.

COMPUTER SKILLS

- Proficient in: Python, Java, C, C++, R, Matlab

- Familiar with: SAS

- SQL, Tableu

PATENTS

- Dynamic Group Creation for Managed Key Servers. Granted: 6/2014

- Sender-specific Counter-Based Anti-Replay for Multicast Traffic. Granted: 6/2011

- Protection of Control Plane Traffic against Replayed and Delayed Packet Attack. Granted: 2/2014

- Updating Machine Learning Models Across Devices. Filed, Number P75341-US01

- Updating Machine Learning Models. Filed, Number P75359-US01

PUBLICATIONS

- Once-for-All Federated Learning: Learning from and Deploying to Heterogeneous Clients. FL4Data-Mining 2023

- Quantifying Catastrophic Forgetting in Continual Federated Learning. ICASSP 2023 · Feb 24, 2023

- Learning from Federated Learning in Real World. ICASSP 2022

- Training Mixed-Domain Translation Models via Federated Learning. NAACL 2022

- PerFedSI: A Framework for Personalized Federated Learning with Side Information. FL NeurIPS 2022

- Communication-Efficient Federated Learning for Neural Machine Translation. NeurIPS 2021 (ENSLP)

- Convergence Analysis of Reweighted Sum-Product Algorithms. International Conference on Acoustic, Speech, and Signal Processing 2007

- **For more of the older publication list, please see**: http://www.frbsf.org/economic-research/economists/tanya-roosta/

AWARDS AND HONORS

- Best student paper award, the International Conference on Acoustic, Speech and Signal Processing, 2007

- National Science Foundation Graduate Student Research Fellowship, 2001-2004

- UC Berkeley Dean's Honor list, 2000

PH.D. SUMMARY

My Ph.D. research was partly funded by DARPA. The overall project was to design and deploy sensor networks for tracking objects and soldiers in the field. One area of research, which was the focus of my work, was understanding various attacks on the software and algorithms used in these networks. Sensor networks have to rely on distributed algorithms for time synchronization and tracking objects. My research work consisted of developing a complete taxonomy of possible attacks, and then focusing on possible statistical solutions to make the algorithms more robust.