

Applied (machine learning) scientist with years of experience in predictive modeling, (sequence-aware) recommender systems, advanced econometric techniques (causal inference from observational data), and A/B testing. Research and publications on personalization, recommender systems, and markets (KDD, WWW, WSDM, Management Science, ISR, POMS, Plos One).

2010-2015 **New York University**. *PhD* in Information Systems (Applied Machine Learning)

2008-2010 University of California Riverside, USA. *MS* in Computer Science

2002-2007 National Technical University of Athens. BS in Computer Engineering (5-year program)

09/2022 - Current *Staff (IC6) Research Data Scientist (Tech Lead)* at **Meta** (formerly Facebook)

- ◇ Machine learning
 - (WIP) Designed and evaluated (A/B testing) predictive frameworks—bootstrapping, probabilistic language models (W2V, LSTM, Transformers), tree-based models (minimum spanning arborescence), and ranking models (Borda count)—for personalized app experience. Likely impact: Improve app experience for billions of users. (Python, R)
 - Developed a framework—PCA, regression analysis—that identifies outlier app experience. Impact: Consistent app experience monitoring that predicts future regressions and guides intervention. Deeper understanding of why such outlier traces occur and how to avoid them. (Python, R, Dataswarm, Unidash)
- ◇ Causal inference and A/B testing
 - Observational analysis—fixed effects, propensity score matching, causalimpact—to robustly estimate wins. Impact: Measurement and trade-off analysis that informs VP-level decision-making. (Python, R, SQL)
 - Identified issues with two different experimentation platforms (selection bias, power analysis) and proposed solutions. Impact unblocked XFNs development workflow; increased confidence in experimental results. (Python, R, SQL)
- ◇ Metrics
 - Lead discussions on efficiency/productivity metric definition. Impact: Designed metrics that inform VP-level decision-making. (Python, R, SQL, Dataswarm, Unidash)
 - (WIP) Lead a strategy-setting project on defining revenue-impact metrics. Impact: Better prioritization of XFN projects through a uniform framework for trade-off analysis. (R, SQL)

07/2015 - 09/2022 *Assistant professor* of Data Analytics at **Boston College**

- ◇ Machine learning
 - Built and evaluated *reinforcement learning* frameworks to provide career path recommendations; proposed frameworks show a 6% increase in market revenue and a 22% increase in worker wages. (Python, R, SQL, [published paper](#))
 - Designed *sequence-aware recommender systems* that match workers with potential employers; predictive performance up to 40% better than state-of-the-art baselines. (Python, R, SQL, [published paper](#))
 - Proposed machine learning frameworks that predict user engagement; defined new engagement metrics that increase predictive performance by up to 40%. (Python, R, SQL, [published paper](#))
 - Designed dynamic expertise assessment systems (HMM, W2V) that yielded 20%–60% better outcomes than state-of-the-art baselines. (Python, R, SQL, [published paper](#))
- ◇ Causal inference
 - Explained (panel data, instrumental variables) the trade-offs of skillset diversification. (Stata, [published paper](#))
 - Explained biases that affect the reputation systems of various online platforms. (Stata, [published paper](#))
 - Explained (natural experiment) the effect of purchase verification on a market's reputation system. ([published paper](#))

06/2015-10/2015 *Machine learning scientist* at **Upwork**

- ◇ Fine-tuned through feature engineering the predictive performance of the platform's employer-worker matching algorithm. Impact: 4% revenue increase and 8% increase in outcomes. (Python, SQL, [published paper](#))

Before 2015 *Research scientist (intern)* at **Microsoft Research**, and *Machine learning scientist (intern)* at **oDesk**.

Topics and methodologies I have used: Predictive modeling, supervised and unsupervised learning, feature engineering, recommender systems, reinforcement learning, hidden Markov models, deep learning, neural networks, text mining, topic modeling, NLP, time-series forecasting; A/B testing, difference-in-differences, panel data, instrumental variables, Heckman two-stage selection models, propensity score matching, survival analysis.

Programming languages and infrastructure I have used: Python, R, SQL, Stata, (before 2015: Java, C#, C++, C); Dataswarm (Airflow equivalent), Presto, MySQL, Manifold, Unidash, Deltoid3, (some experience with FBlearner, MongoDB, portable batch system, Hadoop, MapReduce); VSCode, Rstudio, PyCharm, Vim.