John Urbanik

390 2nd St, Apt 13, Brooklyn NY

(860)-878-1010
johnurbanik@gmail.com
/johnurbanik on github/twitter/linkedin

Mission-driven and ethically conscious engineer and data scientist. Demonstrates strong intuition and is able to dial into details while keeping the big picture in sight. Adept at context switching between research, applied data science, data engineering, people management, and strategy work.

Passionate about the development of end-to-end systems that promote equity or enable new workflows, insights, and experiences. Emphasizes communication, transparency, team, and mission.

EXPERIENCE

otto hearing, Founder, CTO —— Feb 2019 - Mar 2020

- Developed software to improve the listening experience for people with hearing impairments. Built an at home hearing test capable of augmenting standard audiogram with information about sensitivity / dynamic range.
- Utilized hearing test results to adapt the output of a computer in real time to match hearing impairment. Explored methods (ML and classical DSP based) to improve perceptual qualities of music including predictive dynamic range compression, multiband FIR filter design, adaptive Q bandpass filters, and deep differentiable DSP.

Predata, Lead Data Scientist and Data Engineer —— May 2015 - Apr 2018

- Developed regularized hierarchical time-series regression + classification models using techniques from robust/causal inference for prediction of political events from web traffic metadata.
- Developed unsupervised time-series models (anomaly detection, decomposition) utilizing techniques from network theory and knowledge graph engineering in response to business need for interpretable models.
- Built high-cardinality time-series data warehouse with <1s query latency for analytic queries on >500M time series (>1E13 points) per commodity instance.
- Maintained and rearchitected data pipeline, cutting compute cost per model of nightly processing by ~10x.
- Developed processes and pipeline for repeatable, measurable data science experiments, cutting data science and analysis iterations from weeks to days.
- Managed data science/engineering organization, growing it from 1 to 6 people. Championed data quality, EDA, and statistical rigor across the company.

Palantir, Forward Deployed Engineer —— Aug 2014 - May 2015

- Led a team that developed and E2E product that utilized gradient boosting and mixed-integer linear programming to allow non-technical users to specify constraints and preferences interactively to optimize television advertising scheduling to maximize viewership through a user-friendly interface.
- Initiated cultural and process changes through feedback to management. More details on request.

Poptip, Software Engineer (acquired by Palantir) —— Sep 2013 - Aug 2014

- Architected real-time non-parametric Bayesian topic models for detecting changes in themes in Twitter streams.
- Implemented data pipeline running NER, spam detection, and other NLP tasks on up to >100k tweets per minute.

EDUCATION

Princeton University, B.S.E. Electrical Engineering —— 2009-2013

Received certificates (minors) in Applications of Computing and Robotics and Intelligent Systems.
Developed novel approaches for various tasks in speech and audio processing including speaker diarisation, speaker recognition, and blind source separation using graphical models, particle filtering, and various DSP techniques.

Recurse Center, Fellow —— May 2020 - June 2020

Self-directed programming retreat. Continued volunteering on projects related to COVID-19. Mentored other retreat participants on signal processing and machine learning, while self-studying category theory, computational epidemiology, and network theory.

Worked on open source predictive caching project, playlist sharing app, and research into automata with dynamic rules.

CONTRACT / VOLUNTEER WORK

HVN, Technical Advisor —— Apr 2020 - Current

- Helped business make COVID-19 transition as the nature of travel accommodations changed. Developed methodology to dynamically adjust goals and priorities based on incoming travel restrictions and demand flux.

HVN, Senior Software Engineer (Contract) —— Mar 2019 - Apr 2020

- Engineer building data pipelines and conducting spatiotemporal analysis with those pipelines. Also involved in improving search infrastructure and algorithms, performance engineering, and developing engineering culture.

Bellomo Productions, Software Architect —— Jun 2018 - Feb 2020

- Worked with a VFX team as a sole contributor developing software to simulate the physical processes that generate film grain on top of (and reacting to) HD footage in real time. Built a CUDA/C++ based plugin for DaVinci Resolve that gives directors/editors unprecedented perceptual control over digital film.

Princeton University, Research Assistant (Sharad Malik) —— Jun 2011 – Feb 2012

- Developed algorithms for DARPA IRIS program, focusing on automated reverse engineering of logic circuits from gate level to a functional level. Utilized techniques from SAT solving, graph theory, and machine learning.

Beertending, Co-Founder/CTO —— Oct 2012 - Jun 2013

- Built faceted search engine and qualitative recommender system for beer. Developed business model and platform for crowd-sourced craft beer happy hours.

SITU, Engineering Mentor —— Oct 2019 - Apr 2019

- Mentored early career engineers on data engineering and data visualization for human rights related projects.

CoronaWhy, Contributor —— April 2020 - July 2020

- Developed technical spec for machine learning powered literature review for COVID-19 with distributed, volunteer-powered research organization. Spec available on request.
- Worked with computational biology researchers at PNNL on developing methods to extract causal statements from literature and structuring causal inference based methods for understanding viral pathogenesis.

OpenCovid19, Technical Project Lead —— Mar 2020 - July 2020

- Built open science tools for epidemiological modeling. Conducted initial work on an open data exchange, methods for incorporating this data into simulations, and tools for including stochasticity, heterogeneity, and uncertainty in forecasting and health economic planning. See Epicenter for more details.

SKILLS

Languages: Strong in Python, Go, JS, SQL. Working knowledge of Java, Scala, C/C++, R, Julia, MATLAB.

Frameworks and Tools: PostgreSQL, Docker, AWS, Spark, Kafka, Airflow, Pandas, scipy stack, xgboost/LightGBM, TensorFlow/PyTorch, NoSQL (graph, KV, time series, document), Streamlit, ElasticSearch, Parquet, Hudi, k8s

Techniques: Exploratory data analysis, data cleaning, feature engineering, time series analysis, Bayesian statistics, GIS, data flow programming, data warehousing, integer programming, signal processing, algorithmic bias, deep learning, real-time computation, distributed systems, genetic algorithms, multi-task learning.

Soft Skills: Cross-functional communication, business acumen, stakeholder management, team building, data ethics, systems thinking, adaptability.

Technical Interests: Explainable machine learning, human centered computing, causal inference, complexity science, computational neuroscience / biology / epidemiology / sustainability / audio, agent-based models.