

Eric R. Xu, Ph.D.

eric.r.xu@gmail.com | LA & Bay Area | 317.828.9007

I am a results-oriented data professional with a strong background in data engineering, experimental psychology, ad technology, and data product development. I specialize in designing, building, and maintaining scalable data pipelines and systems that handle complex workflows efficiently. My expertise includes data modeling, optimization, and integrating diverse data sources to ensure seamless operations while maintaining data quality, consistency, and reliability. Proficient in Python and SQL, I have a deep understanding of programming and enjoy leveraging task orchestration technologies to enhance workflows and implement best practices for long-term impact. Collaborating with cross-functional teams to deliver data-driven solutions is a key aspect of my work, enabling me to drive meaningful insights, support strategic decision-making, and enhance user satisfaction.

SKILLS

Data Expertise: Engineering | Science | Dashboarding | Modeling | Infrastructure | Governance

Programming Languages: Python | Shell | MATLAB | HTML | JavaScript

Data Orchestration Tools: Apache Airflow | Dataswarm | ETLXpress

Data Technologies: MySQL | Hive | Spark | Presto | ElasticSearch | Snowflake | Apache Iceberg | CockroachDB | Mode Analytics | Google Colab

Dashboarding Tools: Unidash | Looker | Tableau | Sigma Computing

PROFESSIONAL EXPERIENCE

TEKSYSTEMS

August 2024 - present

Analytics Engineering contractor

- **APPLE**
 - Provided engineering support and strategic recommendations to enhance Apple Journey Monitoring across subscription services including Arcade, Music, and Fitness+.
- **DOORDASH**
 - Supported data-driven strategies to enhance the "Dine Out" feature of in restaurant deals, increasing user engagement and restaurant participation
 - Maintaining and updating forecasting models to predict milestones, such as reaching 50K cumulative redemptions, enabling informed product growth decisions.
 - Analyzing A/B test results to evaluate the performance of new features, such as suggested deals and new map vendors, to ensure a safe experience for consumers prior to launch
 - Led the development of OLAP analytics Sigma dashboards to flexibly monitor hourly growth, market expansion, cohort retention, and conversion metrics by different cuts.

INSTAGRAM | META

July 2020 – April 2024

Data Engineer, Analytics

- Designed and implemented robust data pipelines, optimized data models, and leveraged advanced analytics tools to extract actionable insights and drive data-driven decision-making across the organization
- Developed an end to end CPU savings metric using PHP to gamify data pipeline efficiency, boosting CPU saving code changes by 15%, significantly improving overall system efficiency and resource management.
- Provided data engineering support for Instagram's Brand, Identity, and Teen Safety teams, utilizing technologies like Presto and Spark along with star schema data modeling, enabling precise quantification of A/B test results for software enhancements by the teams such as an
 - (a) Increased recovery of benign Instagram accounts by 15% by increasing safety precautions through enhanced user data signals;
 - (b) a 10% reduction in Instagram teen safety violator interactions with minors through interventions such as removing features for accounts affected;

(c) and a 2% reduction in the delivery of risky media content adjacent to ads by enhancing media signal coverage.

INMAR INTELLIGENCE | OWNERIQ

July 2015 – June 2020

Senior Data Engineer | Scientist | Analyst

- Used AI's association rule learning method to develop a data report on the top products viewed and/or purchased together, enhancing customer engagement in analytics offerings.
- Utilized Elasticsearch and the K-nearest neighbors ML algorithm to find lookalike models for incremental return on ad spend estimations, increasing client renewals
- Designed and built effective and scalable web application-based data visualizations for internal and client facing insights across brick and mortar industries with online and offline commerce.
- Led the assessment of feasibility and defined technical requirements of new feature/market requests in partnership with the platform management and client analytics teams.
- Participated in engineering sprint planning to design, optimize, and harden new and existing data product pipelines utilizing MySQL, Presto, Hive, Python, Airflow, Bash, Looker, Tableau, and Spark.
- Developed an end-to-end Python Flask web application to allow various departments to request new beta data products on an ad-hoc basis, resulting in an internal tool for market testing.
- Designed and refined over 8 standalone data products/visualizations, including AB statistical tests to validate the significance of experimental results and predictions.

EPIDEMICO

January 2015 – June 2015

Data Scientist

- Implemented dimensionality reduction of text features in the sentiment model using NLP and AI/ML techniques, improving model efficiency.
- Utilized cross-validation and classification metrics (e.g., precision, recall, AUC, F1 score) to determine the best algorithm for the sentiment model.
- Produced an online API for the developed topic-based sentiment model for automated classification of sentiment in social media posts on e-cigarettes.
- Established a streamlined workflow for auto-categorizing real-time e-cigarette sentiment on social media, enabling curators to provide targeted feedback that continuously improved model accuracy and performance.

EDUCATION

MIT xPRO 6 Month Professional Certificate in Game Design

(Feb 2025 - present) Virtual

MIT Schwarzman College of Computing

Data Science & Machine Learning

(June 2024 - Nov 2024) Virtual

475 Hour Data Science Curriculum Completion

(2014) Metis by Kaplan, New York, NY

Doctor of Philosophy (Ph.D.) Psychological and Brain Sciences

(2008-2014) Dartmouth College, Hanover, NH

Bachelor of Science (B.S.) Biomedical Engineering with Specialty: Biomechanics

(2002-2006) Case Western Reserve University, Cleveland, OH