Jieyi **Deng**

+1 848-252-8487 | dengjy26@gmail.com | portfolio | linkedin | publication

Skills

Programming Skills Python, Hadoop, SAS, Spark, SQL, Tableau, MySQL, Scikit-Learn, TensorFlow, Excel, Matlab, Git, Prompt Engineering

Industrial Knowledge Machine Learning Algorithms, Exploratory Data Analysis, Data Visualization, Statistical Analysis, Hypothesis Testing

Education

Master | University of Washington - Seattle *Washington, USA*

Information Management, Data Science and Artificial Intelligence *Aug 2026*

Master | Rutgers University *New Jersey, USA*

Engineering, Transportation data analysis and risk evaluation *Oct 2018*

Experience

Senior Data Analyst, User Growth | miHoYo | Singapore *Nov 2022 - Mar 2025*

- **Advertising Metrics Monitor:** Designed real-time performance dashboards and formulated a quantitative measurement methodology for SKAdNetwork. Accelerated campaign optimization cycles through improved attribution visibility, driving a 13% uplift in overall mobile campaign ROAS.
- **Anti-Fraud Strategy:** Developed a fraud mitigation framework by analyzing online and offline attribution data to detect suspicious install and activation anomalies. Implemented detection logic that safeguarded campaign integrity and prevented budget waste on invalid traffic.
- **Ad Optimization:** Optimized ad traffic segmentation by cross-classifying device hardware (GPU) and in-app behaviors. Refined the conversion signal loop by suppressing postbacks for low-performing devices, resulting in a 5% uplift in campaign ROI during experiment.
- **Ad Creative Analysis:** Optimized video creative strategy by classifying assets based on granular performance data (CPC, CVR, Purchase Rate). Directed the refinement of ad content based on these insights, resulting in an 8% improvement in Install Conversion Rate.

Data Scientist, Tencent QQ | Tencent | Shenzhen, China *Nov 2021 - Nov 2022*

- **Recommendation System:** Developed a GBT model with an incremental learning architecture to classify high-propensity users based on weekly event roll-ups. Replaced the legacy rule-based system, achieving a 3x increase in CTR and enabling dynamic adaptation to shifting user preferences.
- **User Stratification & Survival Analysis:** Built a hybrid feature matrix combining profile attributes and behavioral time-series to train a Cox Proportional Hazards model for churn prediction. Improved lifetime prediction accuracy by 2.9% and identified strategies to retain 10% of silent users.
- **Data Governance & Pipeline Engineering:** Led a department-wide data governance framework and engineered an automated data quality monitoring system. Streamlined data access and reusability, significantly reducing feature engineering time and enhancing model input reliability.

Data Scientist, Claim Home Office | GEICO | Washington, D.C *Mar 2019 - Sep 2021*

- **Catastrophic Claims Prediction (Geospatial Analytics):** Led a catastrophe analytics initiative by integrating National Hurricane Center (NHC) geospatial tracks with internal policy data using GeoPandas. Developed a two-stage ensemble model (Logistic Regression + XGBoost) to predict city-level claim volumes and probability of loss. Provided early-warning signals that optimized claims staffing and financial reserving.
- **Customer Satisfaction Modeling:** Developed an XGBoost-based driver analysis pipeline to identify key determinants of Net Promoter Score (NPS). Translated feature importance into targeted QA playbooks, driving actionable improvements in customer experience strategies.
- **Data Infrastructure:** Constructed a unified longitudinal event index on Hadoop to track multi-party claim histories. Audited feature counts to resolve data leakage patterns, ensuring robust ETL processes and model integrity

Data Scientist - Intern | Waldron Inc | New York City, NY *Jan 2019 - Mar 2019*

- **Contextual Bandit Recommender:** Deployed a Contextual Multi-Armed Bandit (MAB) engine to dynamically optimize the exploration-exploitation trade-off for online shopping recommendations.
- **NLP Feature Engineering:** Engineered an end-to-end NLP pipeline to structure unstructured product descriptions, creating a unified feature matrix that enhanced downstream ranking model precision.

Projects

Device-Centric Ad Traffic Optimization | miHoYo

- Objective: Reverse-optimize advertising traffic by identifying and suppressing low-performing device segments.
- Methodology: Leveraged historical data (2020-2024) to match and classify activated devices combined with post-activation in-app behavioral data. Analyzed user retention and monetization (payment) performance across device tags to establish traffic priority.
- Implementation: Developed an optimization strategy by suppressing data signal postbacks for identified low-ROI device lists.
- Impact: Achieved a statistically significant 5% uplift in ROI for the experimental group by effectively filtering low-quality traffic sources.

Advertising Anti-Fraud Strategy & Anomaly Detection | miHoYo

- CTIT Analysis: Constructed Click-to-Install Time (CTIT) distributions by linking ad clicks to install events via GAID/IDFV. Detected anomalies by labeling devices with outlier patterns against cohort baselines.
- Traffic Quality Validation: Computed assist rates and click frequency to identify click-flooding and hijacking. Validated Android metadata to detect emulator farms based on abnormal activation velocity and IP clustering.
- Measurement Hardening: Refined offline attribution rules and in-game account matching logic to improve the accuracy of campaign effectiveness measurement.

Incremental Learning for CTR Improvement | Tencent

- Pipeline Architecture: Built a multi-source ingestion flow for ad exposures/conversions and a scheduled feature builder to assemble per-user feature matrices (social behaviors, sensitivity signals).
- Modeling: Trained a Gradient Boosted Trees (GBT) model with an incremental learning architecture. The system performed weekly roll-ups to merge new events without reprocessing full history, enabling rapid iteration.
- Performance: Outperformed the legacy rule-based system with a 3x higher average CTR and surpassed non-incremental ML baselines

Social App User Stratification and Lifetime Analysis | Tencent

- Survival Analysis: Trained a Cox Proportional Hazards model using hybrid features (profile attributes + behavioral time-series) to estimate individual survival curves, improving lifetime prediction accuracy by 2.9%
- Activity Prediction: Constructed lifetime-windowed features and trained a LightGBM model to predict next-day login probability, segmenting users into activity tiers.
- Application: Guided product strategy for target segments, retaining ∼10% of silent users and increasing average session duration by ∼4 minutes via A/B testing.

Catastrophic Claim Prediction for North American Auto Insurance | Geico

- ETL Pipeline: Automated the ingestion of historical hurricane tracks from the NHC, standardized via GeoPandas, and joined with internal policy data.
- Spatiotemporal Features: Engineered features quantifying local storm impact, including center-to-center distances and cumulative exposure hours.
- Ensemble Modeling: Implemented a two-stage approach using Logistic Regression to score loss probability and an XGBoost regressor to estimate total claims volume.
- Deployment: Delivered ZIP-level risk rankings via an internal API to prioritize claims staffing and resource allocation ahead of major weather events.