Anahit Shekikyan

Greater Los Angeles Area | shekikyananna357@gmail.com | 747.305.7037
linkedin.com/in/anna-6a2s44 | github.com/AnahitShekikyan | shekikyan.com

EDUCATION

University of San Diego | M.S. Applied Data Science (GPA: 3.86) 01/2024 - 01/2026
Coursework: Statistics, Data Science Programming, Foundations of Data Science, Machine Learning, Applied Data Science for Business, Applied Data Mining, Applied Predictive Modeling, Applied Time Series Analysis, Cloud Computing (AWS)

Yerevan State University | B.S. Mathematics 09/2002 - 05/2006

CSU Long Beach Digital Skills Bootcamp | Certificate, Data Science and Analytics 07/2022 - 04/2023

PROJECTS

Cybersecurity Capstone (Team Project): Machine Learning for IoT Intrusion Detection (CIC-IoT2023)

- Built a multiclass IoT intrusion detection pipeline using a cleaned/deduplicated CIC-IoT2023 dataset (~21M network-flow records, 105 IoT devices, 33 attacks across 7 families + benign)
- Performed scalable feature engineering in Polars and applied majority undersampling for training while preserving imbalanced validation/test splits
- Trained and tuned 7 supervised models with 5-fold stratified cross-validation; best model was tuned LightGBM (Accuracy 0.7799; Macro F1 0.6329; Weighted F1 0.7963; Micro ROC–AUC 0.9838; Macro ROC–AUC 0.9771)
- Interpreted model behavior using SHAP to identify key predictive signals (e.g., packet rate and TCP-related features) and analyzed tradeoffs for rare attack families
- Selected for publication in the University of San Diego MS–Applied Data Science "Capstone Chronicles" (expected 2026)

Space Waste (Team Project): AWS + ML for Space Debris Risk Analytics

- Built a cloud-based pipeline for five orbital-debris datasets (252,035 records; 83 features) stored in AWS S3 and modeled in SageMaker Studio, including preprocessing and feature standardization
- Evaluated multiple approaches: SageMaker XGBoost for debris growth (93.70% accuracy) and HDBSCAN clustering (33 clusters; 41.74% noise rate)
- Tested collision-risk regression (best RMSE 0.00024, R^2 −0.0026) and documented limitations and recommendations for improved modeling

Classifying Breast Cancer (Python, scikit-learn)

- Built an end-to-end ML pipeline to classify malignant vs. benign tumors using WDBC (n=569) with stratified 5-fold cross-validation and a recall-focused evaluation approach to reduce false negatives.
- Achieved 95.61% accuracy, 95.24% recall, ROC–AUC 0.996 on a held-out test set; delivered a reproducible notebook and technical report on GitHub.

Credit Card Fraud Detection (Team Lead, 3-person team): Imbalanced Classification

- Built a fraud detection workflow on 284,807 transactions (fraud = 0.17%), including duplicate removal (1,081 rows) and robust preprocessing using RobustScaler → StandardScaler to reduce outlier impact
- Led a 3-person team, coordinated tasks, and integrated results; compared supervised and unsupervised approaches using precision–recall curves (AUPRC)
- Best model was Random Forest (AUPRC 0.80) compared with Logistic Regression (AUPRC 0.70)

TECHNICAL SKILLS

- **Programming:** Python, R, SQL (MySQL)
- **Machine Learning:** Classification, Regression, Clustering, PCA, Feature Engineering, Model Selection, Hyperparameter Tuning
- **Model Evaluation:** Cross-Validation, Confusion Matrix, Precision/Recall, F1, ROC–AUC, Threshold Tuning
- **Data Analysis:** Data Wrangling, EDA, Statistical Analysis, Data Cleaning, Preprocessing
- **Visualization:** Tableau, Power BI, Excel
- **Cloud:** AWS (S3, SageMaker)
- **Tools:** VS Code, Jupyter Notebook, Git/GitHub
- **Professional:** Communication, Leadership, Team Collaboration

EXPERIENCE

High School Math, Science, and Computer Science Teacher Arcadia, CA
Arroyo Pacific Academy; Renaissance Academy 10/2021 - Present

- Delivered standards-aligned instruction in Algebra 1–2, Geometry, Precalculus, Chemistry, Physics, Python, and Tableau.
- Created and implemented engaging lesson plans, increasing student comprehension and retention by 20%.
- Integrated data science and analytics into the curriculum, increasing student engagement and performance by 15%.
- Used data analysis to track student progress and differentiated instruction based on performance trends.