# NIKOLAJ KIM

(949)-870-8492 | nikolajkkim@gmail.com | linkedin.com/in/nikolajkim

## EDUCATION

**University of California, Irvine - B.S. Data Science, *GPA: 3.89/4.0*** **04/2026**

**Relevant Coursework**: Machine Learning, Introduction to Artificial Intelligence, Information Visualization, Computational Photography and Vision, Data Structures, Algorithms, Data Management, Statistical Methods for Data Analysis, Probability and Statistics

**Skills**: Python, R, Scikit-learn, SQL, PyTorch, NumPy, Pandas, Jupyter, Tableau, Java, LaTeX, Git, Matplotlib, AWS, Excel, React, C++

**Certifications**: AWS Certified Cloud Practitioner

## EXPERIENCE

**Data Science Research Fellowship** **06/2024 – 08/2024**

*National Science Foundation* *Irvine, CA*

- Developed predictive models for assessing earthquake-induced bridge column damage as part of a 6-member team, utilizing **Exploratory Data Analysis (EDA)** and machine learning techniques, including **LASSO regression**, **Generalized Additive Models (GAMs)**, and **Neural Networks**.
- Achieved **70%** of GAM predictions with an error margin of $\leq 0.05$ and **92%** with $\leq 0.1$, significantly improving accuracy in damage state estimation from simple regression.
- Designed and trained **Neural Networks** that achieved **RMSE** values as low as **0.078** and **R-squared** values up to **0.88**, showcasing robust predictive performance.
- Presented findings at a symposium to 100+ participants, highlighting the models' potential applications in **risk assessment**, **budget allocation**, and **infrastructure safety enhancements**.

## PROJECTS

**AI Golf Caddy** | *Personal Project* **06/2025 - Present**

- Designed and implemented a reinforcement learning–based golf strategy model to recommend optimal clubs, shot types, and aim points based on course layout, hazards, wind, and player tendencies.
- Built a simulation environment from scratch (Python, Shapely, PyTorch), including shot physics, wind sampling, hazard detection, and reinforcement learning agent training.
- Created custom state/action representations incorporating terrain penalties, dispersion models, and hazard proximity encoding.
- Established supporting infrastructure: shot simulator, dynamic action space, state normalization, and visualization tools.

**Capstone Project - Predictive Modeling for Water Leak Risk** | *Team Member* **01/2025 - 06/2025**

- Developed and evaluated machine learning models (**XGBoost**, **LSTM**, **Feedforward Neural Network**) to predict the likelihood of water-related issues (leaks/overflows) in residential units using 38 months of IoT sensor data from 4 properties.
- Addressed severe class imbalance of about 20% positive class using stratified train-test splits, class weighting, and threshold optimization to improve F1 score on minority class.
- Realized highest performance with XGBoost (**F1 = 0.89**, **PR AUC = 0.70** for leaks), enabling risk scoring and real-time integration with Saya Life's smart building platform.

**Autonomous Recruitment Rover** | *AI Developer* **12/2024 – Present**

- Collaborating with a team of 16 engineers to design and build an autonomous rover that engages with students on campus, distributing flyers and initiating conversations about membership opportunities for a campus organization.
- Developing an AI-powered chatbot using **Rasa**, an **NLP** software, to provide personalized event information, aiming to facilitate 1,000+ student interactions and improve engagement rates by **20%**.
- Implementing speech-to-text and text-to-speech capabilities using OpenAI's **Whisper** and **Piper**, enhancing voice recognition accuracy and enabling natural interactions with users.

**Second Place - UCI Datathon** | *Cafe Success Analysis* **04/2025**

- Constructed a predictive pipeline to estimate Yelp ratings for new cafes, empowering owners to optimize operational and customer-facing decisions.
- Used **DistilBERT**, a transformer-based **NLP** model to analyze thousands of Yelp reviews and assign sentiment scores for food quality and service.
- Leveraged **Pandas** for data wrangling and **Seaborn** for exploratory analysis and visualization of trends across cafe attributes.
- Enabled scenario testing for prospective cafe owners by simulating how changes (e.g., reducing weekly hours, improving food sentiment) could influence predicted Yelp ratings.

## EXTRACURRICULARS

**UCI School of Information and Computer Sciences** **01/2024 – 03/2024**

*Lab Tutor* *University of California, Irvine*

- Assisted lab sessions for a Python course, providing guidance to a diverse group of 30-40 students while addressing inquiries in a clear and supportive manner to foster a conducive learning environment.