Matthew Lim

Pleasanton, CA • matthewlim@berkeley.edu • (925) 719-6872 • LinkedIn • matthewlim.me

SKILLS

Languages: C++, Python, Java, SQL, HTML, CSS, C#, JavaScript
Technologies: AWS, NumPy, Pandas, Matplotlib, Scikit-Learn, Flask, Node.js, React, TensorFlow, PyTorch, OpenCV
Tools: Git, Linux/UNIX, Docker, Agile, Unit Testing, VS Code

WORK EXPERIENCE

PricewaterhouseCoopers (PwC) **San Francisco, CA**
Generative AI - Software Engineer Intern *Jun 2024 – Present*
- Developing Azure Functions in C# for PwC's Concourse tool, integrating metadata into Azure Service Bus with HTTP triggers and Postman, and automating client response delivery via Power Automate
- Collaborating with team to design, implement, and test AI-driven features, ensuring continuous improvement and alignment with project goals through scrum's iterative sprints and feedback cycles

Conagra Brands **Chicago, IL**
Generative AI - Technical Project Manager *Feb 2024 – May 2024*
- Engineered a proprietary GPT model using the PrivateGPT API for Conagra Brands, streamlining the ingestion and analysis of investor earnings call data to enhance strategic decision-making processes for both investors and staff
- Leveraged PrivateGPT's customizable GenAI framework, including FastAPI and LLamaIndex, to create a secure, context-aware AI application that supports various local and remote LLM, embeddings, and vector store providers

PricewaterhouseCoopers (PwC) **San Francisco, CA**
CareerVillage - Data Science Consulting Intern *Jun 2023 – Jul 2023*
- Delivered a pro-bono client strategy, ranking top 7% (6/77) of the firm's intern cohorts nationwide
- Analyzed a 37k+ row dataset using Excel, Python (NLP), and Power BI to enable data-driven client strategies

EDUCATION

University of California, Berkeley *Expected Graduation: May 2025*
B.A. Data Science & Business/Industrial Analytics
- **Relevant Coursework:** Data Structures & Algorithms, Principles & Techniques of Data Science, Probability for Data Science, Data Engineering, Object Oriented Programming, Human Ethics of Data
- **Activities:** Data Science Society Project Manager, Theta Tau Professional Engineering Fraternity VP

PROJECT WORK

Matthewlim.me | *AWS (S3, Route 53, CloudFront, DynamoDB, CI/CD), Python, JavaScript* *Jun 2024 – Jul 2024*
- Developed a resume website on AWS using HTML, CSS, and JavaScript, integrated a visitor counter with DynamoDB, API Gateway, and Lambda in Python, and followed best practices in IaC and CI/CD pipelines with GitHub Actions

Music Mate | *Python, Pandas, Seaborn, Scikit Learn, Random Forest* *Feb 2023 – Feb 2024*
- Deployed an interactive Flask website that analyzes user listening habits with Spotify API's insights based on a comprehensive EDA, feature engineering, hyperparameter tuning, and model evaluation
- Trained a mood classification Random Forest model on 1k+ Spotify tracks, analyzing key song features

Smart Environmental Monitoring System | *Python, Pandas, Flask, Plotly, Arduino* *Oct 2023 – Nov 2023*
- Implemented 5 environmental sensors (e.g., AQI, radiation, etc.) via Raspberry Pi on a physical chassis
- Led the backend development of real-time analysis of sensor data in an interactive Plotly dashboard

Spam Email Classifier | *Python, Pandas, Scikit Learn, Logistic Regression, Feature Engineering* *Nov 2023*
- Processed 10k+ labeled/unlabeled samples, engineered features using techniques like one-hot encoding, and implemented logistic regression for classification, achieving a training accuracy of ~85%
- Employed hyperparameter tuning via GridSearchCV, evaluated model performance with metrics (e.g., accuracy, precision, recall, F1, etc.), and visualized the ROC curve for comprehensive assessment

Predicting Housing Prices | *Python, Regex, Scikit Learn, Regularization, Feature Engineering* *Oct 2023*
- Conducted housing price prediction analyses utilizing feature engineering, regression, and CV
- Achieved <200k in testing RMSE and ~90% accuracy across 500,000+ Cook County housing records