Matthew Lim

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

EDUCATION

University of California, Berkeley *Expected Graduation: Fall 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 Co-Ed Fraternity Vice-President

WORK EXPERIENCE

PricewaterhouseCoopers (PwC) **San Francisco, CA**

Generative AI - Software Engineer Intern *Jun 2024 – Present*

- Developed Azure Functions in C# for the Concourse tool, integrating metadata into Azure Service Bus with HTTP triggers, Postman, and automating client response delivery via Power Automate, enhancing efficiency and reducing response time by 40%.
- Engineered Azure Functions in Python for the BYOK tool, utilizing REST APIs (SharePoint API and Microsoft Graph API) to retrieve user metadata, improving file security solutions within the tool's RAG and achieving a 30% increase in data retrieval efficiency.
- Collaborated with teams to design, implement, and test AI-driven features, ensuring continuous improvement with project goals through Scrum's iterative sprints and feedback cycles in Azure DevOps, leading to a 25% increase in project delivery speed.

Conagra Brands **Chicago, IL**

Generative AI – Machine Learning Consultant *Feb 2024 – May 2024*

- Engineered a proprietary GPT model using the PrivateGPT API for Conagra Brands, streamlining the ingestion and analysis of hundreds of investor earnings call data to enhance analysis for investors and staff, resulting in a 31% reduction in analysis time.
- Leveraged a custom Gen AI 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, which improved data handling efficiency by 35%.

PricewaterhouseCoopers (PwC) **San Francisco, CA**

CareerVillage - Data Science Consulting Intern *Jun 2023 – Aug 2023*

- Delivered a pro-bono client strategy, ranking in the top 7% (6/77) of the firm's consulting intern cohorts across the United States.
- Analyzed a 37k+ row dataset using Excel, Python (NLP), and Power BI to enable data-driven client strategies for client success.

SKILLS

Languages: Python, C++, Java, C#, HTML, CSS, SQL, JavaScript
Technologies: Airflow, Spark, Tableau, PowerBI, AWS, NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, TensorFlow, PyTorch, OpenCV
Tools: Git, Docker, Linux/UNIX, Agile, Jira, VS Code

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, ensuring a scalable portfolio.

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, providing users with personalized music listening insights.
- Trained and cross-validated a mood classification Random Forest model on 1k+ Spotify tracks, analyzing key song features to predict mood on user playlists with high accuracy and relevance, achieving an accuracy rate of ~82% and a precision rate of ~80%.

Smart Environmental Monitoring System | *Python, Pandas, Flask, Plotly, Arduino* *Oct 2023 – Nov 2023*

- Implemented 5 environmental sensors via Raspberry Pi on a physical chassis, providing real-time monitoring and data collection.
- Led the backend development of sensor data in an interactive Plotly dashboard, offering users engaging visualizations and graphs.

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% to effectively filter spam emails through key spam words.
- 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, ensuring a robust and reliable spam classification system.

Predicting Housing Prices | *Python, Regex, Scikit Learn, Regularization, Feature Engineering* *Oct 2023*

- Conducted housing price prediction analyses utilizing feature engineering, regression, and CV, achieving <200k in testing RMSE and ~90% accuracy across 500,000+ Cook County housing records, providing accurate and actionable housing market insights.