Matthew Lim

Pleasanton, CA • matthewlim@berkeley.edu • (925) 719-6872 • linkedin.com/in/matthewjlim25 • matthewlim.me

EDUCATION

University of California, Berkeley
B.A. Data Science & Business/Industrial Analytics   **GPA:** 3.7
- **Relevant Coursework:** Data Structures & Algorithms, Object Oriented Programming, Principles & Techniques of Data Science, Probability for Data Science, Data Engineering, Industrial and Commercial Data Systems, Machine Learning & Artificial Intelligence
- **Activities:** Data Science Society Project Manager, Theta Tau Professional Engineering Co-Ed Fraternity Vice-President

WORK EXPERIENCE

Mattel   **Remote**
Data Engineer Contractor   *Sep 2024 – Present*
- Designing and implementing a scalable data pipeline and master data repository to consolidate ESG KPIs across departments into **Google BigQuery**; generate **Tableau** dashboards that benchmark Mattel's metrics against industry standards' sustainability goals.

PricewaterhouseCoopers (PwC)   **San Francisco, CA**
Software Engineer Intern   *Jun 2024 – Aug 2024*
- Attained a 35% boost in operational efficiency and responsiveness by spearheading the design and implementation of **Azure Functions** in **C#** to integrate metadata into **Azure Service Bus**, automating client response delivery through **Power Automate** and utilizing **REST APIs** with **Postman** requests to interact with **Graph APIs** and **SharePoint APIs** for retrieving Microsoft user data.
- Fortified security protocols by developing **Python** functions that rigorously assessed file-level permissions for **SharePoint** files; implemented robust security filters in **Azure AI Search** to protect user data and ensure compliance with regulatory standards.
- Enhanced user experience by revamping the front end for a generative AI application, engineering **Power Fx** within **Canvas** application screens to streamline the application's chat history submissions through intuitive design and improved functionality.

Conagra Brands   **Remote**
Machine Learning Engineer Contractor   *Feb 2024 – May 2024*
- Achieved a 31% reduction in analysis time by engineering a proprietary GPT model utilizing the **PrivateGPT API** for Conagra employees to analyze quarterly investor earnings call data, enabling faster insights through a privacy-focused, localized solution.
- Optimized data handling efficiency by 35% through developing a secure, context-aware AI application using a custom generative AI framework; explored and tested various LLMs, along with technologies such as **FastAPI** and **LlamaIndex**, to optimize performance; and implemented a **Retrieval Augmented Generation** pipeline for efficient document ingestion and contextual responses.

CareerVillage   **San Francisco, CA**
Data Science Consulting Intern   *Jun 2023 – Aug 2023*
- Ranked in the top 7% (6/77) of PwC's consulting intern teams through brainstorming and delivering strategies for CareerVillage.
- Conducted comprehensive analysis of 37k+ rows of user data using **Excel**, **Python**, and **Power BI**, employing **Natural Language Processing** and **Sentiment Analysis** techniques to extract demographic insights and visualizations for customer segmentation.

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%.

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.

SKILLS & INTERESTS

Languages: Python, SQL, Java, C++, C#, HTML, CSS
Technologies: Spark, Tableau, PowerBI, AWS, Azure, Docker, REST APIs, NumPy, Pandas, Matplotlib, Scikit-Learn, Flask, Plotly, TensorFlow, PyTorch, OpenCV, Relational Databases, PostgreSQL, DynamoDB
Tools: Git, Docker, Linux/UNIX, Agile, Confluence, CI/CD, Relational Algebra, Database Normalization, GitHub Actions, Postman
Interests: Cooking, Coffee Brewing, Music, Jazz Alto Saxophone, Hackathons, Gym & Fitness, Bouldering, Reading, Video Games