Ayan Gaur

+1 (858) 281-1254 || ayan.gaur@gmail.com || ayangaur26.github.io || github.com/ayangaur26

EDUCATION

University of California San Diego — Sep 2022 - Mar 2026
B.S. Mathematics and Computer Science

Cornell University — May 2024 - Aug 2024
Machine Learning Foundations Certificate

EXPERIENCE

American Express — Aug 2024 - Dec 2024
ML Intern

- Fine-tuned DeBERTa and DistilBERT models to develop a sensitive data discovery pipeline, achieving 79% accuracy and applying custom regex solutions to accurately identify incorrectly classified data by 30%
- Optimized BIO-label alignment using tokenization strategies, reducing over-redaction and false positives by 25% through cross-validation and hyperparameter tuning, using iterative model optimization
- Integrated contextual NER techniques, achieving 76% precision, 72% recall across diverse financial text datasets

Image Infosystems — Jul 2024 - Sep 2024
Computer Vision Intern

- Developed high-accuracy OCR models with 94.4% accuracy and ¡4.8% loss while using a custom ONNX function converting TensorFlow models to PyTorch, ensuring complete retention of model accuracy
- Improved cheque verification accuracy by 18% through algorithm enhancements for image preprocessing and feature extraction in high-security bank applications

PROJECTS

User Rating Prediction Recommender System Paper *(Python, Pandas, NumPy, Scikit-Learn)* Sep 2024 - Dec 2024
- Conducted EDA on a Food.com dataset (200k+ datapoints), identifying patterns for predictive modeling
- Enhanced scalability by filtering data and applied regularization techniques to reduce model overfitting
- Proposed future enhancements, including hybrid collaborative filtering and deep learning-based embeddings
- Implemented Word2Vec similarity and SVD latent factor models, outperforming baseline models by 18.3%

Computer Vision Personal Project *(Python, OpenCV, Scikit-learn)* Jan 2025 - Mar 2025
- Implemented Bag of Words for image classification using SIFT feature extraction and KMeans clustering
- Applied Bayesian Estimation for email spam classification, leveraging Max Likelihood and Max A Posteriori
- Developed image processing pipelines integrating Edge & Corner Detection and feature-based matching
- Designed and rendered Photometric Stereo models, exploring light-source variations for 3D surface reconstruction
- Trained a FashionMNIST CNN model with backpropagation, achieving 92% classification accuracy and applying transfer learning on STL-10 for feature extraction

F1 Qualifying Performance Analysis - EDA & Statistical Modeling *(Python, Pandas, NumPy, Matplotlib, FastF1 API)*
- Applied statistical modeling and correlation analysis to assess whether high-curvature sectors had a stronger influence on qualifying performance than straight sectors using race data from FastF1 API
- Preprocessed and standardized data by filtering lap times, handling null values, and normalizing sector times
- Visualized sector importance through performance graphs and impact tables, providing data-driven insights

Optimal Coffee Brewing Parameter Recommendation System *(Python, Gemini API, React, Node.js)* Jun 2025 - Jun 20
- Used the Gemini API to recommend optimal coffee brewing parameters based on bean type, grinder, and brew method.
- Created a full-stack website to provide a user-friendly interface for the system.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, Dart, Bash, R, ARM, HTML, CSS, JavaScript, TypeScript
Frameworks: TensorFlow, PyTorch, Keras, Flask, Flutter, Next.js, Tailwind CSS, React, Node.js
Developer Tools: Git, ONNX, Power BI, Firebase, MATLAB, Microsoft Office (Excel, Word, Outlook)
Libraries: Pandas, NumPy, Seaborn, Matplotlib, Scikit-Learn, React, OpenCV