Daniel Kim

San Jose, CA

Email: jobsfordaniel1@gmail.com
LinkedIn: linkedin.com/in/danielkims
GitHub: github.com/danielkimgithub

Professional Summary

Data scientist skilled in machine learning, NLP, and statistical modeling with experience generating predictive insights, cleaning and integrating large datasets, and communicating findings to non-technical audiences. Passionate about applying data science to empower people, improve workplace experience, and modernize analytics workflows.

Education

Master of Science (M.S.) in Data Science, Concentration in Computational Methods

DePaul University – Chicago, IL

- Graduate Honors & Upsilon Pi Epsilon (UPE) Honor Society
- Relevant Coursework: Programming ML Apps, Advanced Machine Learning, Artificial Intelligence II

Bachelor of Art (B.A.) in Economics - Mathematics

Whitman College – Walla Walla, WA

- Relevant Coursework: Calculus III, Linear Algebra, Intermediate Microeconomics, Intermediate Macroeconomics

Skills

- Language & Tools: Python, R, SQL, Jupyter, Excel, CLI, C
- ML & NLP: scikit-learn, PyTorch, TensorFlow/Keras, supervised and unsupervised algorithms, parameter tuning
- Data & Analytics / Visualization: pandas, NumPy, ETL, feature engineering, model evaluation, matplotlib, seaborn

Experience

Higher Ground Youth & Family Services - Anaheim, CA

Data Analyst **Nov. 2024 – present**

- Implemented standardized data workflows and SOPs, collaborating with departmental teams to eliminate redundancies
- Conducted EDA on 4,000+ client records, identifying program trends and informing leadership decisions
- Built automated data workflows and databases, improving cross-department reporting efficiency
- Designed scalable data preprocessing procedures, reducing data inconsistencies by 30%

Omega Accounting Solutions - Irvine, CA

Senior Analyst ← Associate Analyst ←Junior Associate **Feb. 2022 – Feb. 2024**

- Managed 25+ analytical projects, supporting financial and operational decision-making
- Developed client-facing reports summarizing statistical findings and insights for executives and stakeholders
- Analyzed large, confidential business data to validate incentive and compliance claims exceeding $5M

Projects

Bidirectional Encoder Representations from Transformer (BERT) Model Development for Text Analysis **Nov. 2025**

- Fine-tuned BERT with HuggingFace and PyTorch to classify 50,000+ e-commerce product descriptions
- Improved model accuracy from 92% baseline to 97% (macro-F1 = 0.97)
- Implemented preprocessing, tokenization, stratified splitting, and AdamW-optimized GPU training

Predictive ML Modeling of Formula 1 Driver Performance **July 2025**

- Integrated multi-source F1 racing datasets and engineered performance features
- Increased R^2 from 0.45 baseline to 0.60 with tuned Gradient Boosting
- Reduced RMSE from 5.08 to 4.60 through feature selection and hyperparameter tuning

Machine Learning (ML) Model Development for Spam Email Detection **Mar. 2025**

- Processed and vectorized the 517,000-email Enron dataset using TF-IDF
- Achieved 99% accuracy with SVM—12% higher than Doc2Vec-based models
- Evaluated performance with precision, recall, confusion matrices, and ROC curves