Alice Kim   **650.665.1436**   alicek0914@gmail.com

Skills

- SQL, Python, Pandas, Numpy, Pyspark, R, Streamlit, Scikit-learn, PyTorch, TensorFlow, Hugging Face, Lang chain
- ML, Deep Learning, NLP, LLM, AI, Statistical Test: Classification, Regression, Clustering, Time Series, Decision Tree, Random Forests, XGBoost, CNN, RNN,BERT, LSTM, xLSTM, GPT
- AWS S3, Sagemaker, Athena, Redshift, Azure, Jira, Git, Databrick
- Tableau, Power BI

Achievements

- **Hackathon Winner: [RAG Chatbot Development](#) (2024)**
  - Won a company-hosted hackathon in 2024 by developing an advanced RAG Chatbot that effectively utilized the Gemini API and LangChain framework to enhance user interactions and information retrieval.

Project & Competition

- **Team Lead, [Gemini API App Developer Competition](#) (2024)**
  - Led the development of an AI-driven interior design recommendation system, utilizing Python, Flask, Firebase (backend) and JavaScript, HTML, CSS (frontend). Fine-tuned the Gemini API for personalized design analysis and recommendations, leveraging embedding vectors to retrieve and suggest relevant interior design images.

Experience

Southern Gas Company, Atlanta- Data Scientist//Project Lead

Jan 2025 - Current

- Led development of a gas demand prediction model using ensemble model (XGBoost, LightGBM, KNN and Prophet), achieving over 70% error reduction and over 90% variance reduction through data augmentation, advanced feature engineering, and residual analysis, unlocking over $3M annual potential gas control value.
- Developed an Agentic AI system using a DAG-based LangGraph architecture to transcribe customer calls, summarize conversations and analyze sentiment in parallel, generate notes, and evaluate summary quality using LLM-as-a-Judge.
- Built an agentic RAG chatbot by leveraging expertise in prompt engineering, fine-tuning, retrieval techniques with vector databases, and experience in agentic AI with sequential pattern and GenAI evaluation, such as LLM as a Judge.
- Created a proof-of-concept (POC) report and contributed to the end-to-end design, development, and optimization of machine learning and deep learning models to demonstrate feasibility and improve model performance.
- Researched up-to-date methodologies and techniques to develop and optimize the model, and conducted knowledge transfer (KT) sessions to educate coworkers.
- Developed KPI-driven dashboards to translate model insights into actionable recommendations for leadership, enabling data-driven decision-making.

Cox Communications, Atlanta- Data Scientist

March 2022 - September 2024

- Leveraged advanced analytical and statistical techniques, including machine learning methodologies using Random Forest and XGBoost models, to reveal key data patterns, enhancing ticket handling and time management processes.
- Led and managed Spark ETL pipeline maintenance, including troubleshooting and resolving critical issues, using a multi-faceted approach that delivers immediate, temporary, and permanent solutions for comprehensive problem management.
- Developed and deployed comprehensive Tableau and Power BI dashboards, carefully designed to display key performance indicators (KPIs) and conduct detailed statistical analysis using LOD (Level of Details) and DAX (Data Analysis Expressions)

- Resolved 20% data quality issue in the overall data during the Data Transformation process on AWS EMR by modifying the window function sequence and implementing a duplicate removal function in the PySpark script.
- Led a data validation project, building and deploying dashboards that effectively identified and addressed root causes of data quality issues in the workflow.
- Developed a multi-classification and binary image classification model utilizing one of the CNN (Convolutional Neural Network) pre-trained models, EfficientNet, in PyTorch frameworks for my final semester practicum project under the approval of the leadership

The Coca-Cola Company, Atlanta- Data Analyst
July 2019 - March 2022
- Led end-to-end workflow design and project status tracking for data migration projects, ensuring seamless execution and team coordination.
- Skilled at data cleaning, mapping, and profiling to improve data integrity and reduce data redundancy in
  RDBMS by 80%
- Created SQL queries to identify data that are out of sync between tables and resolved data defects
- Established data string cleansing automation and created validation automation by creating Python functions
- Validated data between the multiple systems, generated standard work process reports, and created analytic reports
- Used Informatica, SAP, Thirsty, Salesforce, and Matrix to access databases, identified and resolved linkage issues, and restructured data models

RAPP, San Francisco- Data Analyst
April 2019 - July 2019
- Participated in the Wells Fargo Community Impact Project, which involved big data extraction and management
- Generated data cleansing with a 100% project report completion rate
- Collected and analyzed web content data from potential clients and consumers, and generated complex reports
- Analyzed big data to identify potential issues and errors, and support the redesign and restructure of data models
- Utilized MySQL and advanced Microsoft Excel for data consolidation and collection purposes, and coordinated with the team to generate complex reports

Walmart Global eCommerce, San Bruno- Senior Data Specialist
Dec 2018 - March 2019
- Managed returned product data and other related online information
- Utilized internal application to access database to retrieve and update warehouse data
- Updated product information from different vendors and locations.
- Supported the resolution of issues or errors that occurred through recent data updates or modifications

Education
December 2023 Graduated,
Georgia Institute of Technology, Atlanta - **Master of Science in Analytics**
May 2018 Graduated,
San Francisco State University, San Francisco- **Bachelor of Arts:** Mathematics for Liberal Arts