Kaci Kus

Data Scientist

kacikus@gmail.com | (978) 382-1981 | Arvada, CO 80002

GitHub | LinkedIn | Website

EDUCATION

M.S. Geology (Thesis) – Colorado School of Mines **– 4.0/4.0 GPA** **May 2021**

- **Publication:** Kus, KB, Jobe, ZR, Laugier, F, Walker, W, Sullivan, M. Quantifying the lateral heterogeneity of distal submarine lobe deposits, Point Loma Formation, California: Implications for subsurface lateral facies prediction. *Depositional Rec*. 2021; 00: 1– 30. https://doi.org/10.1002/dep2.169

B.A. Geology & Studio Art – Colby College **– 3.87/4.0 GPA** **May 2018**

- **Publication:** Robert A. Gastaldo, Kaci Kus, Neil Tabor, Johann Neveling; Calcic Vertisols in the upper *Daptocephalus* Assemblage Zone, Balfour Formation, Karoo Basin, South Africa: Implications for Late Permian Climate. *Journal of Sedimentary Research* 2020;; 90 (6): 609–628. doi: https://doi.org/10.2110/jsr.2020.32

SKILLS

Leadership: Technical mentorship, project planning, task distribution, stakeholder presentation, customer communication
Data Mining/ Processing: Web-scraping, missing data imputation, class-imbalance, transformation
Machine Learning (ML): Feature engineering, dimensionality reduction, supervised and unsupervised learning models, predictive modeling, hyper-parameter tuning, convolutional neural networks (CNN), natural language processing (NLP)
Statistical Analysis: Hypothesis testing, Bayesian statistics, parametric & non-parametric techniques, data distributions

EXPERIENCE

Data Scientist, MORSE Corp | Cambridge, MA **June 2021 – present**

- Wrote data pipelines to transform model outputs, compute metrics, and return standardized results to enable faster turn-around of reports to stakeholders.
- Led analysis team of 10 full-time and co-op employees in Agile development process to plan sprints, maintain schedules, and provide overall technical direction for the successful completion of projects.
- Mentored new employees with daily code reviews to balance long-term functionality with short-term needs.
- Contributed to proposals and white papers to bid for government contracts to expand company ML portfolio.

Research Assistant, Colorado School of Mines | Golden, CO **Aug. 2018 – May 2021**

- Developed metrics to quantify and analyze trends from physical measurements, photogrammetry, and geochemical data.
- Increased workflow efficiency by writing Python and MATLAB scripts to streamline data analysis and visualization.
- Designed CNN to classify thin-section mineralogy using fine-grained semantic segmentation.
- Presented research at national conferences and educated non-technical audiences on business impacts of my insights.

Geoscience Intern, Occidental Petroleum | Houston, TX **May 2019 – Aug. 2019**

- Recommended solutions to save company US$1.7 MM per well by analyzing competitor production data.
- Collaborated on multi-disciplinary team of engineers to integrate data from internal and external resources.
- Presented to >30 technical and non-technical stakeholders to communicate business insights in an understandable way.

PROJECTS

Instagram Hashtag Recommender System | *GitHub*: Developed model to recommend optimal hashtags. **Dec. 2020**

- Wrote modular web-scraper to efficiently automate large-scale data collection from Instagram.
- Developed recommender system using image analysis to extract deep features with CNN.
- Optimized recommendations using NLP, Bag of Words, and Skip-gram algorithms, increasing avg. engagement 150%
- Designed and deployed Flask web application in production using AWS EC2 instance to improve accessibility (link)*.

Predictive Modeling of Stock Prices | *GitHub*: Developed models to predict if stock price will rise or fall. **Oct. 2020**

- Wrote API requests to automate data processing pipelines and collect, organize, and update database.
- Parsed and analyzed sentiment of >35,000 news headlines using Natural Language Toolkit Library to improve modeling.
- Trained, tested, and validated supervised learning models (logistic regression, random forest, gradient boosting, SVM) to assess model performance, increasing accuracy to 73%.

Cluster Analysis | *GitHub*: Compared unsupervised learning models to analyze movie ratings and genres. **Sept. 2020**

- Tuned parameters for machine learning algorithms, such as k-means, DBSCAN, gaussian mixture modeling, and hierarchical clustering to determine optimal customer segmentation groups.
- Optimized dimensionality reduction (PCA, t-SNE, and UMAP) to improve data visualization and communicate results.
- Analyzed cluster components to better understand intra-cluster relationships and customer segmentation.

*Web-app only accessible Monday – Tuesday 8:00AM – 5:00PM EST