

MATTHEW MERRILL

DATA SCIENTIST

✉ matt.merrill94@gmail.com 📞 9255973123 📍 San Francisco, CA in matthew-merrill-246a1b55/ 🌐 merrillm1

SUMMARY

Data scientist with a background in physics and math education, experienced in collecting, cleaning and visualizing large data sets for insight. Proficient in developing and deploying machine learning algorithms into web applications using the R Shiny package in an integrated Python / R interface. A former educator able to communicate data stories to drive organizational change using impactful visualization and presentation techniques.

SKILLS

LANGUAGES: Python, SQL, R, Spark, MatLab

DATA COLLECTION: CSV, API, Web Scraping, JSON

DATA VISUALIZATION: Matplotlib, Seaborn, ggplot2, Plotly

MACHINE LEARNING: Classification, Logistic Regression, SVM, Random Forests, Gradient Boosting, Catboosting, Hybrid Recommender Systems, LightFM, Spark ALS, Unsupervised Learning, K-Means Clustering, Agglomerative Clustering

WORKFLOW: Git and Version Control, Jupyter Notebook, Rstudio, Docker Container

EDUCATION

UNIVERSITY OF SAN FRANCISCO

B.S. Physics

M.Ed. Mathematics

- Minors in Astrophysics and Mathematics.

Sept. 2007 - May 2015

SPRINGBOARD DATA SCIENCE CAREER PROGRAM

Certificate

- Completed 600+ hours of hands-on curriculum, with 1:1 industry expert mentor oversight, and completion of 2 in-depth projects. Mastered skills in Python, R, SQL, data analysis, data visualization, hypothesis testing, and machine learning.

Nov. 2019 - Oct. 2020

PROJECTS

PREDICTING HOTEL CANCELLATIONS

- Achieved a 90% mean AUROC for classifying cancellations with an 83% accuracy rate, improved from 76% and 70%, respectively.
- Identified 10 key indicators of cancellation risk using Catboosts gradient boosting algorithm with SHAP for variable interpretability.
- Developed an interactive web application with the Shiny R package to directly deploy final model into production.
- Value of work: Increase accuracy of projections for daily actualized income, reduce cancellations among higher risk bookings, optimize the distribution of customer channels.

OLIST RECOMMENDER SYSTEM

- Constructed a hybrid recommendation system using data from ecommerce site Olist and established a 97% AUROC using LightFM's hybrid algorithm.
- Attained an 18% increase in AUC using clustering techniques to add additional features to the user features sparse matrix.
- Developed a Spark ALS recommendation algorithm for comparison, accomplished a 0.259 RMSE using cross-validation for parameter tuning.
- Built and deployed docker container using Kubernetes KNative manifest for full on serverless machine learning deployment.

WORK EXPERIENCE

SAN FRANCISCO UNIFIED SCHOOL DISTRICT, Mathematics Educator

Aug. 2015 - July 2020

- Led as department chair over the '17-'18 and '18-'19 school years, overseeing an increase in statewide testing results of 8% average per year.
- Designed study in conjunction with other departments to test grading policy changes with randomly sampled groups, study found an 11% average increase in assessment scores relative to the control.
- Conducted 4 semesters of school wide PD sessions with a team of colleagues, communicating monthly results of research and experimental data collection.
- Drove changes to 5 levels of curriculum structure that aim to close the nearly 30% achievement gap between mainstream and marginalized subgroups.

SEO SCHOLARS, SAN FRANCISCO, Mathematics Instructor

Aug. 2012 - Current

- Taught as an original member of a 10 person team dedicated to building an educational start-up to assist low income and first generation students to go to college.

RELEVANT EARLY EXPERIENCE

ASTROPHYSICS DEPT., UC BERKELEY, Astrophysics Researcher

May 2009 - June 2010

- Cleaned radio interferometric data to construct, display and analyze astronomical images of possible protostar to determine an accurate classification (binary or singular).
- VLA 7mm continuum map revealed a single compact source of emission, not meeting the approximate 0.08 sun mass required to meet assumptions of a separate optically thin disk emission.
- Prepared and presented conclusions at the summer 2010 American Astronomical Society Meeting (AAS).