

# MATTHEW MERRILL

## DATA SCIENTIST

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

## SUMMARY

Data scientist with a background in physics and math education, skilled in creating prediction algorithms for user and customer insight. Proficient in developing and deploying machine learning models into web applications using the R Shiny and Streamlit libraries in an integrated Python / R interface. I am a former educator able to communicate data stories to drive organizational change using impactful visualization and presentation techniques.

## SKILLS

**PROGRAMMING LANGUAGES:** Python, SQL, R, Spark, MatLab

**DATA COLLECTION:** Database query, CSV, API, Web Scraping, JSON

**DATA VISUALIZATION:** Matplotlib, Seaborn, ggplot2, Plotly, Tableau

**MACHINE LEARNING:** Classification, Logistic Regression, SVM, Decision Trees, Gradient Boosting, Catboosting, Hybrid Recommender Systems, LightFM, Spark ALS, Unsupervised Learning, KMeans Clustering, Agglomerative Clustering

**WORKFLOW:** Git and Version Control, Jupyter Notebook, Rstudio, Docker Container, Distributed Computing with Spark

## EDUCATION

### UNIVERSITY OF SAN FRANCISCO

B.S. Physics

M.Ed. Mathematics

- Minors in Astrophysics and Mathematics.

September 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 queries, data analysis, data visualization, hypothesis testing, bayesian inferential statistics, and machine learning.

November 2019 - October 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, modeling with a Catboosts gradient boosting algorithm with SHAP for variable interpretability.

- Developed an interactive web dashboard with the Streamlit library to directly deploy final model into production.

- Value of work: Increase accuracy of business prediction for daily actualized income, reduce cancellations among higher risk bookings, optimize the distribution of customer marketing.

### OLIST RECOMMENDER SYSTEM

- Constructed hybrid recommendation system to increase customer retention 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 customer segmentation labels 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

August 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

August 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).

- Statistical analysis of 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).