MATTHEW MERRILL
DATA SCIENTIST

✉ matt.merrill94@gmail.com 🌐 merrillm1.github.io/ 📞 (415) 335-9957 📍 San Francisco, CA in matthew-merrill-data-scientist/ ⌂ merrillm1/

SUMMARY

Data scientist with experience building robust statistical models to solve common business problems such as customer acquisition, cancellation risk and customer retention using supervised and unsupervised learning methods. Proficient in the development and deployment of machine learning models using R Shiny and Streamlit libraries in an integrated Python / R interface. I am a former educator able to communicate technical concepts to all levels of personnel to drive organizational change.

SKILLS

PROGRAMMING LANGUAGES: Python, SQL, R, MatLab, Excel
DATA SCIENCE LIBRARIES: Pandas, NumPy, Scipy, Matplotlib, Seaborn, Scikit-Learn, Statsmodels, Fastai, Pytorch, BeautifulSoup
WORKFLOW: Jupyter Notebook, Google Colab, Git and Version Control, Tableau
MACHINE LEARNING: SUPERVISED - Logistic Regression, Linear Regression, LASSO, Ridge, SVM, Decision Trees, Gradient Boost, Catboost,
UNSUPERVISED - KMeans, Hierarchical, Agglomerative, RECOMMENDATION - LightFM, Spark ALS, NLP - ULMFiT
MODEL DEPLOYMENT: Streamlit, R Shiny, Docker Container, Binder

EDUCATION

DATA SCIENCE CAREER TRACK, SPRINGBOARD
Certificate Data Science 2019-2020
Completed 600+ hours of hands-on curriculum, with 1:1 industry expert mentor oversight, completion of 2 in-depth projects, and three exploratory data analysis projects.
Mastered skills in Python, R, MySQL, data at scale, data wrangling, data analysis, time-series analysis, data visualization, inferential statistics, hypothesis testing, data modeling, machine learning and data storytelling.

UNIVERSITY OF SAN FRANCISCO
Masters - Math Education 2014 - 2015
Bachelors - Physics, Math minor 2007 - 2012
Learned technical concepts in math and physics and communicated them at a high level to students, peers and professors through colloquiums, research opportunities and lab instruction.

EXPERIENCE

HOTEL CANCELLATION PREDICTION ALGORITHM
Hypothesized source of recent cancellation increases in the hotel industry and gathered data to verify assumptions with predictive analytics.
- Constructed a viable solution as measured by a 0.90 AUROC for classifying cancellations, by using Random Forest and Catboost models for speed and interpretability.
- Optimized model output using robust techniques for feature selection, feature engineering and hyperparameter tuning.
- Quantified feature impact as measured by the SHAP importance metric, by eliminating sources of data leakage and identifying consequential thresholds.
- Constructed an interactive web analytics dashboard with Streamlit to display the affect of marketing channel distribution on revenue and cancellation rate.

ECOMMERCE RECOMMENDATION SYSTEM
Translated a major business need, as identified in a low 4% customer retention rate for an ecommerce website, and provided quantitative solution using machine learning.
- Developed solution by constructing an effective product recommendation system as measured by a 0.97 AUROC using LightFMs hybrid algorithm.
- Achieved an 18% increase in AUC using clustering techniques to add customer segmentation labels to the user features sparse matrix.
- Created a Spark ALS recommendation algorithm for comparison, accomplishing a 0.259 RMSE using cross-validation for parameter tuning.
- Built a docker container using a base LightFM image for machine learning deployment.

PRIMARY FACTORS LEADING TO USER ADOPTION
Identified the most important factors contributing to user adoption for an online product.
- Achieved a 0.81 F1 score, chosen as the primary metric due to the imbalanced nature of the data.
- Manually programmed classification label to reflect users that logged into the product on three seperate days in at least one seven-day period.
- Utilized an SVM model for its ability to manage the size and sparcity of the data as well as for interpretability.
- Prepared and delivered report in a jupyter notebook, refactored with production level code.

EMPLOYMENT

HIGH SCHOOL MATHEMATICS EDUCATOR, San Francisco Unified School District   August 2015 - July 2020
- Led as department chair, overseeing a small team of educators, responsible for conducting weekly meetings, reporting to administration and district officials about departmental progress and needs.
- Organized yearly analytical study to test grading policy changes, communicating findings to staff and influencing key educational stakeholders.
- Drove changes to school-wide curriculum structure by organizing and presenting pedagogical research to all levels of staff as part of a professional development team over a two year period.
- Guided as a charismatic and detail-oriented educator, organizing daily lesson plans with provisional data, communicating technical concepts clearly and effectively to cohorts of 120+ students.

STEM COORDINATOR, Tenderloin Achievement Group   August 2012 - July 2014
- Designed, revised and enhanced daily curriculum for groups of 20+ after school STEM participants grades K-8.
- Managed small programming staff, responsible for organizing projects, learning activities and field trips.
- Assisted in strategizing and researching new funding sources to support educational, nutritional and public programs.
- Wrote grant proposals to foundations, corporations, and government agencies to fund STEM activities.

UNDERGRADUATE ASTROPHYSICS RESEARCHER, Astrophysics Dept., UC Berkeley   May 2009 - June 2010
- Cleaned radio interferometric datasets to construct, display and analyze astronomical images of possible protostar to determine an accurate classification (binary or singular).
- Performed statistical analysis of VLA 7mm continuum map data and revealed a single compact source of emission, not meeting the approximate 0.08 sun mass required to meet assumptions of a seperated optically thin disk emission.
- Prepared and presented conclusions at the summer 2010 American Astronomical Society Meeting (AAS).