MATTHEW MERRILL

DATA SCIENTIST

matt.merrill94@gmail.com | /merrillm1.github.io | (415) 335-9957 | San Francisco, CA | /matthew-merrill-data-scientist | /merrillm1

SUMMARY

Data scientist with end-to-end experience building and deploying robust statistical models into production with real customer data. Recently led a small team of data scientists using agile software methodology to query, clean and mine data for insight and prediction and reported findings to stakeholders weekly. Former mathematics educator and astrophysics researcher, uniquely skilled in communicating technical concepts to all levels of personnel to drive organizational change.

SKILLS

PROGRAMMING LANGUAGES: Python, SQL, R, Spark Parallel Computing, MatLab, Excel
DATA SCIENCE LIBRARIES: Pandas, NumPy, SciPy, Matplotlib, Seaborn, Plotly, iPlot, SciKit-Learn, Statsmodels, Fastai, Pytorch, BeautifulSoup
WORKFLOW: Jupyter Notebook, Google Colab, Git and Version Control, Spark Clusters with DataBricks, Tableau for Data Visualization
MACHINE LEARNING: SUPERVISED - Logistic, Linear, LASSO and Ridge Regression, SVM, Random Forests, Gradient Boosting (XGboost, Catboost)
UNSUPERVISED CLUSTERING - KMeans, KNN, Hierarchical, DBSCAN, Agglomerative RECOMMENDATION SYSTEMS - LightFM, Spark ALS.
MODEL DEPLOYMENT: Streamlit, R Shiny, Docker

EDUCATION

DATA SCIENCE CAREER PROGRAM, SPRINGBOARD

Certificate - Data Science — JAN 2020 - DEC 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 — JUNE 2014 - AUG 2015

Bachelors - Physics, Math and Astrophysics minor — SEPT 2007 - MAY 2012

Learned technical concepts in math and physics and communicated ideas at a high level to all levels of personnel through colloquiums, research opportunities and lab instruction.

DATA SCIENCE EXPERIENCE

MACHINE LEARNING FELLOW, Fellowship.ai — JAN 2021 - PRESENT

Interned full-time for 4 months, spearheading data science projects with a team of fellows using real and recent customer and advertising data.

- Implemented agile software methods to solve problems in campaign image optimization, churn prediction and customer segmentation for Dockers.
- Lead as scrum master for a team of 10 data scientists, reporting daily progress and presenting weekly slide decks to the founder.
- Created data pipelines to extract and transform 100+ features for improving CTOR prediction using image filtering, text extraction and contour detection.
- Improved CTOR prediction accuracy by over 30% compared to baseline models and customer age prediction by 20% using oversampling methods.
- Finalized Streamlit app for campaign CTOR prediction and slide decks for customer segmentation and reported results to key stakeholders.

INDEPENDENT PROJECTS

HOTEL CANCELLATION PREDICTION AND STREAMLIT DEPLOYMENT

Programmed an interactive web dashboard that allows users to alter customer market distribution to lower cancellation rates by upwards of 30%.

- Mined data for influential customer attributes and intervention points to increase daily revenue, potentially reducing the impact of cancellations by 10%.
- Optimized model output using robust techniques for feature selection, feature engineering and hyperparameter tuning, achieving an accuracy of 88%.
- Quantified feature impact by eliminating sources of data leakage and by finding consequential thresholds with importance plots (SHAP).
- Constructed a web analytics dashboard to interact with key findings from 4 consequential attributes of customer history.

E-COMMERCE RECOMMENDATION SYSTEM AND DOCKER DEPLOYMENT

Translated a business need, as identified in a low 4% customer retention rate for an ecommerce website, and provided a 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 by segmenting customers based on purchasing habits and applying this to the user matrix.
- Trained 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.

EMPLOYMENT

MATH DEPARTMENT CHAIR AND EDUCATOR, San Francisco Unified School District — AUG 2015 - AUG 2020

- Conducted weekly meetings to analyze data and discuss interventions, reported to staff, administration and district officials regularly about goals, progress and needs.
- Collected and analyzed data and discussed findings with staff as part of a yearly ROCI (Result Oriented Cycle of Inquiry) cycle.
- Guided as a charismatic educator, organizing daily lesson plans with provisional data and communicating technical concepts to over 120+ students.

ASTROPHYSICS RESEARCHER, Astrophysics Department, UC Berkeley — MAY 2009 - MAY2010

- Cleaned radio interferometric data to construct, display and analyze images of a 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.
- Prepared and presented conclusions at the summer 2010 American Astronomical Society Meeting (AAS).