ROBIN STOPA | DATA SCIENTIST

rstopa901@gmail.com | (508) 207-5409 | San Mateo, CA
linkedin.com/in/robin-stopa | github.com/rstopa901

I am a data scientist with a background in math and data analytics. My previous experience in big healthcare data and consulting helped me develop strong technical skills in python, SQL, and machine learning. I bring empathy and a desire to learn to all teams and projects that I am a part of.

SKILLS

Machine Learning: Linear Regression, Logistic Regression, Bagging, Random Forest, ExtraTrees, KNN, Boosting, Ensembling/Stacking, PCA, K-means Clustering/DBSCAN, Neural Nets/Deep Learning, NLP, A/B testing, naive bayes
Programming Languages: Python (pandas, sklearn, statsmodels, keras, tensorflow, matplotlib, seaborn, sqlalchemy, nltk, spacy, requests, beautifulsoup, pickle, json, pipeline, fuzzywuzzy), SQL, R, html
Software: Pycharm, JupyterLab, PostgreSQL, DBeaver, Git, Github, Git Bash, SVN, JIRA, Tableau, Spotfire, AWS, LaTeX, MS Excel
Languages: Japanese (fluent)

EXPERIENCE

Data Science Fellow, General Assembly 10/2022 - 02/2023
Successfully completed 500+ hours of expert-led instruction in statistics, data management, machine learning, and the industry's most in-demand technologies

Capstone: A study of SMOTE pre-processing techniques in small imbalanced data

- Studied effect of SMOTE pre-processing to solve for target imbalance in small data for both regression and classification. For regression, created a linear regression model on SMOGN pre-processed data to predict the incidence of cheating/hacking in games based on anti-cheat software and game description (NLP). For classification, used video game sales data and created logistic regression model to predict if a video game would generate over 10M sales based on various features of the game. Evaluated model results on regular vs. SMOTE pre-processed data.

Project for Aviation Safety

- Designed logistic regression model for plane safety regulators to predict the severity of an aviation accident based on contextual factors such as the model and make of the plane, and identify the most important factors as possible targets of regulation

Comparing Language in FPS Games

- Created Logistic Regression model to classify reddit posts as *Call of Duty: Modern Warfare II* subreddit or *Overwatch 2* subreddit posts using NLP. Gathered data using Pushshift Reddit API and compared performance of classification models in Python. Determined that *Overwatch 2* subreddit uses more niche gaming vocabulary than *Call of Duty: Modern Warfare II* subreddit. Free-to-play games such as *Overwatch 2* may benefit from a niche language base as this may drive loyal, existing players to make in-game purchases

CONSULTANT, IQVIA - 159 SOLUTIONS 02/2022 - 10/2022

- Led onshore and offshore teams to conduct complex analytics on big healthcare data in the areas of patient analytics, physician targeting, SalesForce optimization and competitive intelligence. Translated complex analytics into actionable insights and recommendations for clients
- Managed all client communication, including presentations to non-technical stakeholders
- Oversaw effort to design a web-based app for all clients as a new product offering

SENIOR ASSOCIATE, AXTRIA INC. 09/2020 - 02/2022

Database Creation for Oncology Treatment

- Utilized SQL to create business rules and designed KPIs for analytic layers of database
- Managed data science team and coordinated with engineering team to deliver analytics-on-demand database
- Led client to purchase database product for two additional drugs
- Engineered analytics-ready databases as standard company offering, which were sold to four additional clients over 12 months

Patient Journey Analysis for COVID-19 Treatment

- Delivered actionable insights on medical treatment journey of COVID-19 patients via automated reporting, enabling leadership team to make fast business decisions
- Strengthened client relationship by tactfully managing expectations and timelines

SUPERVISOR, ANALYTICS PRODUCTS, VEEVA CROSSIX 07/2017 - 01/2020

Over the course of 3 years at Veeva, I was promoted from Associate to Lead Associate, then from Lead Associate to Supervisor, with a focus on the projects below:

Digital Marketing Impact Product

- Revolutionized algorithmic attribution methodology for digital marketing analysis by employing Shapley Values. Productized and automated analysis using python, SQL, and AWS resulting in quadrupling the revenue of this highest-selling analytics product after two years
- Directed company-wide implementation of algorithm by working across client service, engineering, and leadership teams
- Pioneered expansion of game-theory based algorithmic model to TV, cross-channel, and physician analyses by utilizing SQL, python, and AWS

Physician Conversion Analyses

- Acted as BI + technical owner for fourth highest-selling analytics product. Advised users on best practices for analytic methods. Engineered product improvements, especially in automation

Control Variable Research

- Reduced time and complexity for control selection in analytic offerings

EDUCATION

Data Science Certificate, General Assembly 10/2022 - 02/2023

Haverford College, B.S., Mathematics major, Chemistry minor 08/2013 - 05/2017

Thesis: Logistic Regression Model to Determine the Probability of Winning a Game of Tennis