

Daniel Rossetti

Greater Boston | drossetti12@gmail.com | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

DATA SCIENTIST

I'm a data scientist who enjoys building machine learning models that serve as impactful business tools. As a former mechanical engineer, I bring a rigorous problem-solving approach and the ability to bridge theoretical concepts with real-world applications. With these skills, I plan to develop practical solutions with positive impacts.

TECHNICAL SKILLS & CORE COMPETENCIES

PROGRAMMING LANGUAGES & TOOLS: Python, Jupyter Notebook, SQL, Markdown, Git

LIBRARIES: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Tensorflow, Beautiful Soup, Natural Lang. Toolkit

MACHINE LEARNING: Clustering, Classification, Linear & Logistic Regression, Neural Networks, Natural Language Processing, Web Scraping

PROJECTS

- **Boston Airbnb Fair Pricing Tool:** Developed a pricing regression model to estimate Boston Airbnb listing prices. Used web scraping, APIs, and natural language processing (NLP) techniques to engineer features. Implemented a recommender system using cosine similarity to suggest listings based on user inputs. The pricing model outperformed the null model by reducing the root mean squared error from \$86 to \$49 and achieved an R-Squared value of 0.68 on validation data with outliers removed.
- **Alzheimer's Severity MRI Image Classification:** Built an image classification model employing both convolutional and pre-trained neural network models to identify the severity of Alzheimer's disease in MRI images. The model achieved an accuracy of 71% across four severity classes. Also developed a binary classification model focusing on only the two least-severe classes achieving an accuracy of 81%.
- **Identifying Humor With Language Models:** Created classification models employing natural language processing to categorize news article titles collected from the r/worldnews and r/TheOnion subreddits using the Pushshift API. Regular expressions were used to clean the data, removing errors and subreddit identifiers in the titles. A variety of classification models were attempted and evaluated. The best-performing model was a stacking classifier with an accuracy of 86.0% on unseen data.

RELEVANT EXPERIENCE

THE OPPORTUNITY PROJECT (TOP), DATA SCIENTIST – Volunteer [May 2023 – Present]

- Developing a framework for data encoding and medical reporting system integration in a cross-functional team of UX designers, software engineers, and other data scientists tasked with creating an application to report at-home covid test results. TOP is administered by the U.S. Census Bureau and the FDA in 2023.

DRAPER LABORATORY, MECHANICAL ENGINEER – Cambridge, MA [Jan 2016 – Mar 2023]

- Led the mechanical design of a benchtop-scale, multi-chambered dialysis cartridge which outperformed traditional dialysis cartridges with a filtrate fraction approximately 3 times higher than conventional units.
- Designed and built electronics enclosure assemblies for sensitive electronics. Guided printed circuit board (PCB) layout/design to accommodate heat dissipation and electrical/optical/assembly interfaces.

EDUCATION

GENERAL ASSEMBLY, DATA SCIENCE IMMERSIVE – Remote [March-June 2023]
12-week, full-time, 480+ hour program teaching data analysis, data science, and machine learning.

MIT APPLIED DATA SCIENCE PROGRAM – Remote [May – August 2021]

UNIVERSITY OF MASSACHUSETTS LOWELL – Lowell, MA [Sep 2006 – May 2010]
Bachelor of Science in Mechanical Engineering (BSME) – Highest GPA in ME GPA: 3.9/4.0