A.J. Goldsman

linkedin.com/in/aj-goldsman | ajgoldsman@gmail.com | 818-800-2809 | ajgoldsman.github.io/My-Portfolio/

Education & Certifications

Thinkful, Online

- Data Science Flex Program, Certificate of Completion   January 2020
  - **Relevant Coursework:** Programming in Python, Data Visualization & Experimentation, Supervised/Unsupervised Machine Learning, SQL Fundamentals, Statistics & Probability

University of California, Los Angeles, Los Angeles, CA

- B.S. Chemical & Biomolecular Engineering, Minor in Theatre   June 2019
  - **Relevant Coursework:** Computer-Aided Design and Analysis with Pro/II, Computer Programming with C++, Computer Programming with MATLAB, Introduction to Geographic Information Systems

Work Experience

Spatial Front Inc., *Data Scientist*   February 2020 – Present

U.S. Census Bureau, Economy-Wide Statistics Division (Contract)

- Developed a Python application to effectively address over 70% of the million non-useful entries reported with each Economic Census cycle. With the planned integration of this tool into the 2022 Economic Census, we will be able to quantify up to 20% more revenue data than in previous cycles, accounting for billions of dollars in additional revenue.
- Currently developing a Tableau-esque Python application that will allow Census analysts to pull relevant data and produce general visualizations for the purpose of pinpointing areas of focus. This application currently runs as a Flask app and produces visualizations via Plotly Dash.
- Reviewed old Python code written by members of the team that have since left the Census. Translated the Python scripts to more easily understandable pseudocode and generated reports summarizing the programs' functionality, thus enabling current and future team members to better understand and use sections of the code in related projects.
- Worked on several projects in which Supervised Machine Learning was used to develop classification models and interpolation pipelines for matching publicly available data and imputing missing records in Census and IRS databases..

Dr. Playlist, *Co-Founder*   May 2019 – January 2020

- Developed a machine learning pipeline with the goal of improving music classification, with a focus on music genre/style, via analysis of instrument type, chord progressions, meter, etc.
- Built and compared more than a dozen supervised machine learning models to classify over 100,000 tracks from the Free Music Archive within 16 genre labels. Achieved a model testing accuracy of over 90%.
- Developed a Heroku-hosted, browser-based Flask app that recommends Disney songs to a user based on a comparison of various features extracted via SpotiPy API and NLP analysis of the songs' lyrics.

iCardio.AI, *Data Science Consultant*   April 2019 – June 2019

- Designed and implemented a Python application to search for and extract variables from over 50,0000 PDF and Microsoft Word documents to a more useful Pandas DataFrame, accomplished with Python's Textract and PyPDF2 libraries, in addition to Natural Language Processing via the use of regular expressions.
- Analyzed extracted data, using Matplotlib and Seaborn, to determine correlations between variables such as heart valve dimensions and ejection fraction for patients who underwent a Transthoracic Echocardiogram.

NovellusDX, *Data Science Intern*   Summer 2018

- Designed and implemented Python functions utilizing NumPy and Pandas to compare feature data collected by the image recognition software searching plates of cells for specific, introduced mutations.
- Utilized logistic regression, principal component analysis, and t-SNE analysis, as well as other statistical models, to analyze the relationship between the dozens of features from the cells being analyzed and the algorithm's probability of successfully predicting whether a specified mutation was present.
- Built a Python program to automate the data analysis process by determining the most significant experimental features and their appropriate ranges, improving the algorithm's accuracy while reducing the amount of employee time required to tune it.

Skills

Programming Languages: Python, SQL, Javascript, HTML/CSS, MATLAB, C++
Libraries and Frameworks: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, SpaCy, ScraPy, NLTK, Bokeh, Flask
Software Tools: QGIS, Microsoft Office Suite, Linux/ Red Hat, Jenkins, Altair PBS Professional, Adobe Creative Suite