GABRIEL ALON
(650) 799-9295 | gabrielalon257@gmail.com | California
www.linkedin.com/in/gabrielalon https://github.com/galonpy https://galonpy.github.io

Data Scientist with 5 years of experience in Python from work, research and a Masters in Data Science

EDUCATION

University of Michigan, Ann Arbor   May 2022 – August 2023
Master of Applied Data Science

- Created a BERT and ChatGPT-3.5 (OpenAI) recommender system for YouTube videos on a Streamlit web application

- Doubled translation performance with a Hugging Face Transformer LLM on an English to French dataset (Bleu Score)

University of California, San Diego   Graduated 2019
Bachelor of Science in Management Science

PROFESSIONAL EXPERIENCE

McD Tech Labs (Purchased by IBM) | Mountain View, California   June 2019 – July 2021
Senior Data Scientist

McD Tech Labs developed a neural network language model to automate restaurant drive-through ordering with Siri like interactions.

- Programmed statistical models and visualizations with Python and SQL to oversee the performance of the AI on millions of customer interactions, and categorize bugs. Used pandas, scikit-learn, matplotlib, tableau and superset for models and visuals.

- Directly improved millions of customer interactions by fixing machine learning evaluation code. Fixed python bugs in a logistic regression classifier's accuracy metrics which were predicting that regular customer speech was noise.

- Coded an outlier monitoring report for a vision neural network used for detecting customer readiness, via a Python api to Amazon Redshift. The report summarized thousands of weekly interactions though the command line.

- Developed a weekly technical performance pdf used for determining engineering priorities. This report highlighted product trends, specific bug failures, and proposed product requirements to patch bugs and handle unforeseen customer behaviors.

- Developed a manual annotation framework for specifying which neural network or software layer was likely responsible for an AI failure with a customer. This involved extensive collaboration with engineers and reading of academic research.

- Wrote product requirements in Jira for engineers based on the manual evaluation of thousands of customer interactions. Worked closely with product management to develop a realistic and nuanced plan for the product.

- Trained new data scientists and engineers as the company transitioned from 12 people, to 70+ people.

Pure Storage | Mountain View, CA   June 2018 – September 2018
Marketing Data Analyst (Internship)

- Developed Python machine learning models and Tableau reports on web marketing to sales conversion data.

- Created new analytics metrics and doubled attributed prospect profiles across databases, which were adopted by the Sales and Marketing departments.

- Analyzed the search term to site content conversion process using Adobe Analytics on the company website

DealSumm | Palo Alto, CA   June 2017 – September 2017
Marketing Research Analyst (Internship)

- Analyzed the Hubspot email marketing and sales practices of a machine learning lease contracts platform using Excel

- Performed product marketing, customer outreach and demos to firms

Model N | Redwood City, CA   June 2016 – August 2016
Marketing Data Analyst (Internship)

- Created a competitive landscape report of the top 50 Pharma companies used to advise the CEO and the Board prior to the acquisition of their main competitor Revitas.

- Developed a software product marketing spreadsheet with Salesforce data as well as interviews with management and consultant market research reports.

Wefi | Tel Aviv, Israel and Santa Cruz, California   July 2015 – January 2016
Data Analyst (Internship then Part-Time)

- Created SQL, Tableau, and Excel reports on broad mobile app usage patterns from a dataset of 1 million Android phones.

- Published on Venturebeat.com in an article titled "Attention Shoppers: Shopping with a store app means you'll visit more and stay longer" on July 31, 2015.

- Created reports by custom request, ie: informing investors prior to Square's IPO, or analyzing the impact of Clash of Clans' Super Bowl ads on app usage by specific user demographics.

RESEARCH

Neural Network Researcher   July 2021 – June 2022

- Published an NLP research paper analyzing transformer neural net prediction confidence https://arxiv.org/pdf/2206.14348.pdf

- Presented research to 50 A.I professionals in an event organized by the Stanford A.I Professional Program.

Microsoft InterpretML   Oct 2018 – March 2020
Open Source Python Contributor

- Found inconsistencies between the Python feature importance graphs of the GA2M model and the expectations of the equations in the research paper, which invalidated its statistical significance, and led to a correction being issued.

- Discovered a mathematical inconsistency in the examples section of the paper "Purifying Interaction Effects" (Lengerich et. Al) which led to a correction being issued.

SKILLS

Languages: Python (5 years) (Object Oriented, Data Structures, Algorithms, Production Code, Unit Tests),
         SQL (3 years), Spark SQL, PostgreSQL, R, Unix, Bash, Shell, Git

Tools: Jupyter Notebook, GitHub, AWS, Linux Terminal, Tableau, Superset, Excel, Amazon Redshift, Amazon S3, Jira

Libraries: Pandas, Numpy, Scikit-learn, Matplotlib, Pytorch, Hugging Face, Altair, NLTK, TensorFlow, Statsmodels, Seaborn

Models: Machine learning models, Forecasting models, Statistics, Visualizations, Data Science models, Causal Inference, A/B tests, Neural Networks, Transformers, Decision Science, Random Forest, Unsupervised Learning, Clustering, Regression, Deep Learning, Supervised Learning, ETL, Large Language Models (LLM), APIs, GBM, XGBoost, Generative Adversarial Networks (GANS), Network Modeling, Bayesian Models, Explainable A.I., SHAP, Plotly Dash

Additional: CI/CD, Docker/Containerization, REST API, Version Control, Linear Algebra, Microservices, Healthcare data