DAVID B. GOMEZ

Data Scientist / Machine Learning Engineer

✉ dbgomez94@gmail.com
📞 (478) 747-6519
📍 Boston, MA
🌐 https://dbgomez94.github.io/

ABOUT ME

Former engineer turned data scientist (DS) / machine learning engineer (MLE) specializing in social media data and mental health outcomes, especially suicidality. Seeking a role as a DS / MLE for social good.

EDUCATION

Master's of Science
Computer Science - Machine Learning
Georgia Institute of Technology

📅 Aug 2022 - current
📍 Atlanta, GA
🎓 3.8/4.0

Relevant courses
- Deep Learning (A)
- Machine Learning (A)
- Artificial Intelligence (A)
- Social Computing (A)
- Data and Visual Analytics (A)

Mater's of Science
Aerospace Engineering
Georgia Institute of Technology

📅 Aug 2018 - Aug 2020
📍 Atlanta, GA
🎓 3.7/4.0

Awards
- NASA Space Technology Fellowship

Relevant courses
- Accident Causation & System Safety (A)

Bachelor's of Science
Georgia Institute of Technology

📅 Aug 2015 - Dec 2018
📍 Atlanta, GA
🎓 GPA: 3.6/4.0

Awards
- Outstanding Undergraduate Researcher

SKILLS

Python; NumPy; Pandas; Seaborn; Scikit-learn; Keras; Tensorflow; PyTorch; Hugging Face; GitHub; MATLAB; LaTeX; STATA; R; LabVIEW

RELEVANT PROJECTS

Patterns of Suicide Disclosures on Twitter
Graduate Researcher | Social Dynamics and Well-Being Lab

📅 Fall 2023
- Combined machine learning, deep learning, and natural language processing techniques and a sample of 1500 Twitter timelines to identify temporal-linguistic response patterns to suicide-related disclosures on Twitter.
- Built an array of language-based text-classifiers ranging from bi-gram logistic models to pre-trained transformers from HuggingFace to preempt which response type a user would exhibit using historical linguistic data with a F1-score of 0.64.
- Collected data using Twitter's APIs, extracted linguistic features via LIWC lexicon, and developed models in Python using Sci-Kit Learn, Keras, and Hugging Face.

Language Characteristics of Suicide-Related Sub-Reddits
Group Project Leader | Deep Learning Course Project

📅 Fall 2023
- Utilized natural language processing and transformer-based deep learning models to identify language characteristics that differentiate suicide-related forums on Reddit.
- Collected data using Reddit's Pushshift API and developed models in Python using Keras, Tensorflow, and Hugging Face.

Improving the Gun Ownership Proxy for Firearm Research
Graduate Researcher | Space Systems Design Lab

📅 Fall 2020
- Utilized linear regression and deep neural networks to develop an improved proxy for gun ownership for use in firearm-related research that increased the accuracy of the state-of-the-art proxy from $R2 = 82\%$ to 95%.
- Collected data from CDC BRFSS and WONDER databases and conducted analyses in STATA and Python (PyTorch).

Challenging the Inevitability of Suicide
Group Project Leader | Social Computing Course Project

📅 Fall 2019
- Utilized statistical learning and model diagnostics techniques to argue that methods to reduce firearm suicide rates (gun laws) do not exacerbate suicides by other means.
- Collected suicide data from CDC WONDER APIs and conducted analyses in STATA.

WORK EXPERIENCE

Research & Development Engineer
Busek Space Propulsion and Systems

📅 August 2020 - August 2021   📍 Natick, MA
- Performed tests, experiments, and data analyses for the qualification of electric propulsion devices.
- Built LabVIEW applications to automate the acquisition of Hall thruster magnetic field profiles and oscilloscope measurements that reduced acquisition time by 25%.
- Developed a standardized uncertainty analysis procedure in MATLAB for electric propulsion thrust stands.

Electric Propulsion Intern (x3)
NASA Jet Propulsion Laboratory

📅 Summers 2018, 2019, 2020   📍 Pasadena, CA
- [2020] Developed a comprehensive uncertainty analysis procedure for electric propulsion thrust stands using linear regression and inverse prediction intervals.
- [2018, 2019] Built a custom LabVIEW application for electric thruster testing that synchronized the control and data acquisition of 20 devices and over 200 channels and improved the hardware response time and sampling rate by 1 and 3 orders of magnitude, respectively.