Sejal Dua

503-505-0595 | sejal.dua@tufts.edu | sejal-dua-761371164
sejaldua | sejaldua.com

Education

Tufts University — Medford, MA
B.S. Data Science, B.S. Biomedical Engineering | GPA: 3.8 (Dean's List) — *Sep 2017 – May 2021*

Jesuit High School — Portland, OR
GPA: 4.26 weighted, 3.96 unweighted — *Sep 2013 – Jun 2017*

Experience

IBM Research — Jun 2020 – Sep 2020
Machine Learning Intern — *Yorktown Heights, NY*

- Engineered primary NLP classification model for IBM Drug Repurposing for Cancer pipeline by performing corpus filtering through heuristic rules encapsulating domain expertise.
- Achieved 83% accuracy after training a Snorkel-based distantly supervised binary classifier on 127,000 unlabeled PubMed articles and validating on 1,400 labeled articles.

Textbook Exchange Network — Jan 2019 – Present
Director of Data Analytics — *Medford, MA*

- Gathered data-driven insights from 6000+ API exchanges representing textbook transactions that have saved students $500,000 compared to campus bookstore prices.
- Calculated Key Performance Indicators (KPIs) via SQL queries and Python statistical methods to measure the health and wealth of the organization and market estimated impact.
- Managed a team of computer science students through numerous cycles of exploratory data analysis, feature integration, visualization, and presentation at monthly showcases.

LEGO Education Engineering Lab — Jun 2019 – Aug 2019
Software Development Intern — *Boston, MA*

- Developed a Python software package which enables bidirectional communication between a LEGO SPIKE Prime and existing microprocessors, streamlining the creation of extension kits.
- Improved LEGO image recognition functionality by implementing a smile detection neural network demonstration using the sklearn package.

Oregon Center for Aging & Technology — May 2018 – Aug 2018
Software Engineering Intern — *Portland, OR*

- Facilitated and executed a research validation experiment designed to obtain gait speed data from an elderly cohort of participants with Alzheimer's or dementia; engineered laser trip beams and 3D-printed sensor enclosures for passive infrared sensing virtual hallway.
- Wrangled and visualized 20 years of passive infrared sensing data, investigating the gradual degradation of the circadian rhythm as a result of aging.

Projects

TechTogether Boston 2020: Pilter AI | *Awards: IBM Best Hack & Dell Technologies Best Hack* — Jan 2020

- Built an NLP-powered, user-facing annotation tool that sifts through abstracts from therapeutic cancer intervention studies and extracts relevant data using named entity recognition (NER).

Beyond the Lyrics | *Published in Towards Data Science* — Nov 2019

- Performed sentiment analysis on song lyrics using the Spotify API, Python, and Tableau; Wrote 6 Medium articles reaching over 26K readers in the data journalism space.

Technical Skills

Languages: Python, Java, C/C++, SQL (Postgres), JavaScript, HTML/CSS, R, MATLAB, PHP, Bash
Frameworks: React, Node.js, Flask, Tableau, MongoDB, Django
Developer Tools: Git, Jupyter Notebooks, Docker, Google Cloud Platform, Visual Studio Code, Xcode, PyCharm, IntelliJ
Libraries: matplotlib, numpy, pandas, sklearn, scipy, spacy, D3, Keras, Tensorflow, PyTorch, Snorkel