

# SEJAL DUA

Tufts University Class of 2021  
Data Science & Biomedical Engineering

## EXPERIENCE

### IBM Research

#### MACHINE LEARNING INTERN

- 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.

, Yorktown Heights, NY

Jun 2020 - Sep 2020

### Textbook Exchange Network

#### DIRECTOR OF DATA ANALYTICS

- 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, integration, visualization, and presentation at monthly showcases.

Medford, MA

Jan 2019 - Present

### LEGO Education Engineering Lab

#### SOFTWARE DEVELOPMENT INTERN

- 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.

Boston, MA

Jun 2019 - Aug 2019

### Oregon Center for Aging & Technology

#### SOFTWARE ENGINEERING INTERN

- 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.

Portland, OR

May 2018 - Aug 2018

## CO-CURRICULARS

### Tufts University Computer Science Department

#### TEACHING FELLOW

Jan 2019 - Present

Hosted office hours, facilitated labs, and improved grading infrastructure for 120+ person classes; Served as liaison between professor and 30+ person teaching staff.

### Tufts Women in Technology Conference

#### SPEAKER COORDINATOR

Apr 2020 - Present

Networked with women in executive positions at top tier tech companies; Recruited 8 speakers to inspire and empower college students at virtual women in technology conference.

### Towards Data Science

#### FREELANCE DATA JOURNALIST

Nov 2019 - Present

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.

## PROJECTS

### TechTogether Boston 2020

#### 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)

### URx InternHacks

#### AWARD: BEST TECHNICAL PROBLEM

Jun 2020

Integrated a Flask web application that facilitates intellectual conversation between users with a K-Means Clustering algorithm, to help mitigate social media echo chambers.

## CONTACT



(503) 505-0595



sejal.dua@tufts.edu



sejaldua



www.sejaldua.com



sejal-dua-761371164

## EDUCATION

### Tufts University (Medford, MA)

GPA: 3.77 Dean's List

Graduation: May 2021 (expected)

### Jesuit High School (Portland, OR)

GPA: 4.26 weighted, 3.96 unweighted

Graduation: June 2017

## RELEVANT COURSES

- Data Structures
- Algorithms
- Machine Learning
- Data Visualization
- Neural Networks and Deep Learning
- Big Data
- Artificial Intelligence
- Cybersecurity

## TECHNICAL SKILLS

### Coding Languages

Python, C, C++, Java, R, SQL, HTML, CSS, Javascript, PHP, Matlab, Bash

### Frameworks & Technologies

React, Node.js, Flask, Tableau, MongoDB, Postgres, Django, Docker, Makefile, Git

### Packages & Libraries

matplotlib, numpy, scipy, spacy, pandas, sklearn, D3, Keras, Tensorflow, PyTorch, Snorkel

### Productivity Tools

VSCode, Jupyter Notebooks, Xcode, LaTeX, Microsoft Office Suite, Trello

### Microprocessors & Fabrication

Raspberry Pi, Arduino, micro:bit, OpenMV Camera, PyBoard, SolidWorks, 3D Printing, Laser Cutting

## INTERESTS

Fitness Tech, Music, Chinese, Sports Analytics, Sushi, Travel, Documentaries, Yoga & Meditation, Data Journalism