Sejal Dua

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

Education

Tufts University — Medford, MA
B.S. Data Science, B.S. Biomedical Engineering | GPA: 3.82 (Summa Cum Laude) — *Sep 2017 – May 2021*

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

Experience

Nike - North America Nike Direct Stores Analytics — Dec 2021 – Present
Data Analytic Manager — *Beaverton, OR*

- Engineered a pipeline to collate a temporally dynamic 100+ feature dataset at the store grain; developed a stakeholder-facing Streamlit store clustering web application using K-Means Clustering, Gaussian Mixture Models, and Principal Feature Analysis to support merch, planning, and allocation decision-making (85% adoption rate)
- Created a "Win With Sport" Tableau dashboard which was socialized throughout NA planning and assortment teams with the objective of achieving optimal gear up and licensed product assortment coverage and productivity across key sports, leagues, and teams in stores; implemented flexible, weighted ranking methodology for aggregating multiple signals
- Launched a post-store opening analysis framework (SQL, Python, Tableau) with the objective of empowering the Unite Concept team with early insights for new store openings throughout NA stores acceleration; leveraged Google's Causal Impact Analysis model in order to quantify amount of business created and/or cannibalized by new store opening

IBM Watson Health — Jun 2021 – Dec 2021
Data Scientist — *Cambridge, MA*

- Implemented client-facing predictive model (3x better than baseline) for Medicaid enrollment forecasting; identified most optimal model across 5+ exogenous and non-exogenous time series forecasting approaches including Holt-Winters exponential smoothing, ARIMA, SARIMAX, and Keras models.
- Collaborated with Program Integrity analysts to develop parameterized PySpark algorithms for detection and further investigation of fraud, waste, and abuse patterns across providers.
- Leveraged Streamlit and Networkx libraries to build a dashboard illuminating patient-provider trends within MarketScan de-identified claims database, as discovered via graph algorithms.

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 – Jan 2021
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.

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 12 Medium articles reaching over 60K readers in the data journalism space.

Technical Skills

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