

Ash Jha

650 278 6911 | San Francisco, CA | ashwinijha28@gmail.com | ash-jha.github.io | linkedin.com/in/ashjha1

Education

University of San Francisco | Master of Science in Data Science June 2020 (Expected)

- Courses: Machine Learning, Deep Learning, Experiment Design, A/B Testing, Distributed Computing, Time Series Analysis, Data Visualization, Data Structures & Algorithms

University of Toronto | Bachelor of Applied Science in Industrial Engineering May 2015

- Courses: Operations Research, Probability & Statistics, Calculus, Linear Algebra

Experience

Data Scientist, Intern October 2019 - Present

Retrace Labs | San Francisco, CA

- Identified potentially fraudulent claims by developing a Neural Network in PyTorch to detect similar claims. The model architecture was made part of a patent that is currently under review.
- Built a document classifier using PyTorch that automatically detects the type of form a customer is submitting, bypassing the need for manual selection.
- Decreased run time of image processing code from 7 days to 5 hours by writing more efficient MongoDB queries and parallelizing Python code using multiprocessing.

Senior Data Analyst September 2015 – May 2019

Canadian Tire | Toronto, Canada

- Saved \$300k in transportation costs by creating a linear programming model in Python to determine the most cost-effective mode to deliver products to stores.
 - Responsible for the end-to-end process including creating the data cleaning scripts, formulating the model and creating reporting for end-users.
- Predicted damage rate for new products by creating a Linear Regression Pipeline in scikit-learn.
- Decreased product spoilage rate by mapping transportation routes against external temperatures in Tableau. Gathered and cleaned the data using SQL and pandas.

Business Operations Analyst, Intern May 2013 – September 2014

IBM | Toronto, Canada

- Developed “people analytics” dashboards for senior management using Cognos, SQL & Excel.

Projects

Outbreak - Interviewed by YCombinator December 2019 – March 2020

- Predicted flu outbreaks using social media data. Interviewed by the non-profit arm of YCombinator.
- Created features from Instagram posts (hashtags, image content) and historical flu trend data.
- Achieved a 91% R^2 by feeding these features in to a Random Forest model and performing hyperparameter tuning using randomized grid search.

Kaggle Competition - Bengali.AI (Top 10%) December 2019 – March 2020

- Created computer vision models using PyTorch to classify components of 200k+ images of handwritten Bengali characters. Received a Bronze medal for placing in the Top 10% of competitors.

Machine Learning Algorithms from Scratch November 2019

- Coded my own implementations of Regression (with L1/L2 penalties), Decision Trees, Random Forest, Feature Importance and k-means algorithms from scratch in Python.

Software & Technologies

Python (pandas, NumPy, scikit-learn, PyTorch, matplotlib, Seaborn, Bokeh, Flask) | R | SQL | Tableau | Git | Jupyter | Apache Spark (PySpark, SparkSQL) | MongoDB | AWS (EC2, S3) | Google Cloud