

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

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

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

- Courses: Operations Research, Calculus, Linear Algebra, Object Oriented Programming, Fundamentals of Accounting & Finance, Markets & Competitive Strategy

Experience

Data Scientist October 2019 - Present
Retrace Labs | San Francisco, CA

3rd data science hire, responsible for developing machine learning models to automate the dental claim adjudication process. Retrace is funded by Intel Capital, Softbank Ventures and Stonepoint Capital.

- Worked closely with full-stack and data science team to deploy several machine learning models to production using Python, AWS, Flask, Unicorn, nginx and Docker
- Wrote a CNN + LSTM model for written handwritten character recognition, allowing parsing of thousands of documents that open source packages were unable to accurately parse.
- Improved accuracy of dental claim form matching by 2% by using TFIDF on tri-grams of document id strings and using k-nearest neighbors to find the closest match.

Senior Data Analyst September 2015 – May 2019
Canadian Tire | Toronto, Canada

- Saved \$300k in transportation costs by creating a linear programming optimization 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 visualizations for end-users.
- Decreased product spoilage rate by mapping transportation routes against external temperatures in Tableau. Gathered and cleaned the data using SQL and pandas.

Projects

Outbreak - Interviewed by YCombinator December 2019 – March 2020

- Predicted flu outbreaks using social media data. Interviewed by the non-profit arm of YCombinator.
- Achieved a 91% R^2 by running features from social media data and historical flu data through a Random Forest model and performing hyperparameter tuning using randomized grid search.

Kaggle – Ranked in top 2.5% of all competitors worldwide December 2019 – Present

- **Bengali.AI (Bronze Medal - Top 10%)** - Classified components of 200k+ images of Bengali characters using VGG and resnet deep learning networks in PyTorch.
- **Melanoma Classification (Bronze Medal - Top 10%)** - Classified whether a patient had melanoma or not by combining medical images and tabular data using EfficientNet ensembles in PyTorch.

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 | GitHub | Jupyter | Apache Spark (PySpark, SparkSQL) | MongoDB | AWS (EC2, S3) | Google Cloud | Docker