

ADITYA VYAS

adityavyas17@gmail.com • <https://adityavyas17.com> • <https://github.com/aditya1702> • <https://kaggle.com/aditya1702>

EDUCATION

School of Arts and Sciences, Rutgers University

Aug 2018 - May 2020

Master of Science in Data Science, GPA: 4.0/4.0

Dhirubhai Ambani Institute of Information and Communication Technology

July 2014 - May 2018

B. Tech. (ICT Honors with Minors in Computational Science), GPA: 7.6/10

RELEVANT COURSEWORK

Data Structures and Algorithms, Machine Learning, Introduction to Artificial Intelligence, Probability and Statistics, Massive Data Storage and Retrieval, Discrete Mathematics, Data Interaction and Visualisation, Database Management System, Operating Systems

INTERESTS AND EXPERTISE

Research Interests: Machine Learning, Data Science, Algorithm Trading, Quantitative Finance, Portfolio Optimisation

Tools: Scikit Learn, Pandas, Numpy, ggplot2, Pytorch, Tensorflow, Apache Spark, OpenAI Gym, OpenCV, DeepDive, Flask, Django

Programming Languages: Python, R, Java, SQL, Javascript, Powershell, CSS, HTML

OPEN SOURCE CONTRIBUTIONS

mlfinlab, *Co-Author and Member*

July 2019 - Present

Scikit-Learn, *Contributor*

March 2019 - Present

RESEARCH

Hudson & Thames Quantitative Research, *Quant Researcher*

July 2019 - Present

- Collaborating with a team of 6 researchers to implement academic research in buy-side investment management.
- Implemented different portfolio optimisation algorithms - Hierarchical Risk Parity, Critical Line Algorithm and Efficient Frontier Methods - as part of an open source Python package - mlfinlab.

INDUSTRY EXPERIENCE

Apteo, *Data Science Research Intern (co-op)*

Sept 2019 - Dec 2019

- Used alternative data to forecast quarterly financial metrics - revenue, gross margin, earnings per share - for different companies.
- Increased forecast accuracy by 5% using feature engineering techniques - upsampling, lagged features and percentage change features.

Barclays, *Software Developer Intern*

June 2019 - Aug 2019

- Helped in transitioning over to a new monitoring system - ITRS Geneos - for equities trading applications.
- Automated the setup of a test infrastructure of 9 servers with log-files and processes mimicking production environment.
- Wrote automation scripts to convert old system configuration files to new one and speed up the transition process.

Apteo, *Data Science Intern*

Aug 2017 - Aug 2018

- Created a Deep Q-Network parametric reinforcement learning agent to execute trades for around 6000 stocks each day.
- Replaced word vectors (word2vec) with document vectors (doc2vec) in existing neural network model. Reduced mean absolute error (MAE) of the model by 3%.

Shipmnts, *Machine Learning Intern*

May 2017 - July 2017

- Created a novel algorithm to detect repeated structures in an image with 80% accuracy using Fuzzy similarities and other visual and semantic heuristics.
- Constructed features and models to detect table headers in images and pdf documents. Final Random Forest model achieved an 82% recall using techniques like undersampling, oversampling, and SMOTE analysis.

KAGGLE COMPETITIONS

- Two Sigma: Using News to Predict Sentiments: Top 18%
- Petfinder.my Adoption Prediction: Top 10%
- Santander Value Prediction Challenge: Top 9%
- Costa Rican Household Poverty Prediction: Top 18%
- Quora Insincere Classification Challenge: Top 12%
- Dont Overfit-II: Top 4%

ACADEMIC PROJECTS

Create a Reddit Comment Generator using BERT and GPT2

Sept 2019 - Oct 2019

- Trained XGB, Neural Networks and DistilBERT models on 15GB of reddit comments for subreddit classification.
- Used the classification BERT model and reddit data to finetune OpenAI's GPT2 model for reddit comment generation.
- Deployed the final models as a web API using Docker and Google Cloud - <https://aditya1702.github.io>

ManyLabs Data Imputation and Interpolation

March 2019 - April 2019

- Implemented a machine learning pipeline from scratch (using only Numpy) to clean and analyse the psychology study dataset.
- Tested Numpy implemented models for imputation - SVM, Linear, Ridge, Lasso, Random Forests, Decision Trees & Mixture Models

Music Genre Belief Recognition

Sept 2018 - Dec 2018

- Recognised music genre from audio samples using convolutional neural network with time-distributed layers.
- Improved the out-of-sample accuracy on the FMA dataset from 50% to 60%.