DASHANSH PRAJAPATI

dp3085@columbia.edu | (646) 409-3080 | www.linkedin.com/in/dashansh

Education

Columbia University — New York, NY
Master of Science in Data Science, GPA: 3.7/4 — Jan 2021 – Dec 2022
Coursework: Applied Deep Learning, Applied Machine Learning, Probability and Statistics, Data Analysis & Visualization.
E-board: President, Data Science Institute Student Council.

Indus University — Gujarat, IN
Bachelor of Technology in Computer Engineering, GPA: 9.61/10 — Aug 2016 – Jun 2020
Coursework: Big Data Analytics, Python Programming, Data Warehouse and Mining, Analysis of Algorithms.

Skills

Programming	Python, R, SQL, C, C++.
Packages	Numpy, Pandas, Scipy, Tensorflow, Scikit-learn, Keras, OpenCV.
Data Visualization	Tableau, Power BI, Data Studio, Looker, ggplot2, matplotlib.
ML Techniques	Classification, Regression, Clustering, Statistical Modeling, Bagging, Boosting, XGBoost.

Experience

Omdena — Remote
Machine Learning Engineer Intern — May 2021 – July 2021
- Executed synonym replacement and language translation to upsample minority class size by approximately 20%.
- Spearheaded a team of 35 annotators to label more than 20k text samples for 6 weeks (Tool: Labelbox, Python).
- Trained Multinomial Naive Bayes, SVM, Logistic Regression and BERT classifiers with maximum accuracy of 83%.
- Developed a pipeline employing scikit-learn, to extract Twitter and Reddit data and predict probability of 7 classes.

SvaKatha — Gujarat, IN
Data Science Intern — Jun 2020 – Sep 2020
- Analyzed 10k+ fashion portraits to obtain trending color palettes belonging to 8 different categories using OpenCV.
- Reduced 75% of preprocessing time of background removal by executing a deeplab-v2 model of 80% mIoU score.
- Designed a closet recommendation engine to suggest top 5 matching clothes using color and structure of cloth.

Indian Space Research Organization (ISRO) — Gujarat, IN
Deep Learning Research Intern — Dec 2019 – May 2020
- Built an image processing tool to preprocess any high-resolution image to generate 100X low resolution samples.
- Evaluated performance of various semantic segmentation architectures: DeepLab, Unet, and SegNet.
- Automated architecture search process by generalizing hyperparameter search using KerasTuner.
- Proposed a variation of Unet architecture; achieved 5% increase in mIoU score on the ISPRS's Potsdam dataset.

Project

MAP48 Challenge, Morgridge Family Foundation (1st place) — Oct 2021 – Dec 2021
- Developed an R package extending ggplot2 functions to make visualizations customized to client's branding guide.
- Documented the package and deployed it on Github; Placed 1st in a 48-hour hackathon.

Recommender System for Video Games — Oct 2021 – Dec 2021
- Implemented Content-Based Filtering and Collaborative Filtering Approaches to recommend items to users.
- Estimated effect of sentiment scores of reviews, metadata and summary of reviews on item recommendations.
- Achieved 16.4% and 17% overlap between recommended items and also view and also buy items respectively.

Abstractive Text Summarization — Jan 2021 – Apr 2021
- Trained a seq2seq model with attention to summarize 500,000 samples from amazon's fine food reviews dataset.
- Surveyed text classification models of varying complexity: LSTM, Stacked LSTM and Attention Mechanism.