VINAY KALE

180 Claremont Avenue, Apt.65, New York, NY 10027
vk2392@columbia.edu | (646) 510-1024 | vinaykale64.github.io

EDUCATION

Columbia University New York, NY
M.S. in Data Science, GPA: 3.76/4 Expected Dec 2018

- Relevant Coursework: Probability Theory, Statistical Inference and Modeling, Machine Learning, Natural Language Processing, Neural Network & Deep Learning, Algorithms, Computer Systems

Indian Institute of Technology Madras Chennai, INDIA
B.Tech. in Mechanical Engineering, M.Tech. in Product Design, GPA: 8.12/10 June 2016

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Hive, Caffe, Java, C, C++, MATLAB, LaTeX, MapReduce, Spark
Tools and Technologies: Tensorflow, PyTorch, Keras, Git, Tableau, RShiny, Amazon EC2, OpenCV, CUDA, MS Office

PROFESSIONAL EXPERIENCE

COLUMBIA UNIVERSITY New York, USA
Graduate Course Assistant Aug 2018– Dec 2018

- Teaching Assistant for COMS 4996 Applied Deep Learning under Prof. Joshua Gordon
- Coursework supervision including fundamentals (DNNs, CNNs, RNNs), advanced topics like GANS, NMT and deployment applications on mobile and web

SPOTIFY New York, USA
Data Science Intern (Product Insights Team) June 2018–Aug 2018

- Developed a fan-artist pair segmentation pipeline which quantifies affinity of any user towards an artist, covering over 16B user-artist pairs.
- Analyzed how a new feature affects a typical fan-artist journey and how different fan segments engage with it, by deploying appropriate AB tests and conducting statistical significance tests **[BigQuery, Python]**

COLUMBIA UNIVERSITY New York, USA
Graduate Course Assistant Jan 2018–May 2018

- Teaching Assistant for COMS 4995 Applied Machine Learning under Prof. Andreas Mueller
- Coursework supervision including Git, Continuous Integration (Travis, Jenkins), Scikit-Learn and Keras modules

ZS ASSOCIATES Pune, INDIA
Data Scientist June 2016–July 2017

- Built a novice AI Doctor using patient vector embeddings (similar to Word2Vec) trained on sequential big medical data. This addresses multiple questions in PLD (patient level data) space where patient level classification in different contexts is often sought after [**Python, SQL, Tensorflow, AWS**]
- Created a marketing-mix based spend analyzer data pipeline which calculates impacts of different marketing campaigns employed using regression based models with elastic net regularization **[R, SQL, Tableau]**

PUBLICATION

CVPR Conference 2018: Traffic Surveillance Research Workshop Paper New York, NY
Research under Prof. Zoran Kostic Sep 2017–April 2018

- Used Mask-RCNN for object detection and localization, Deep-Sort for object tracking with goal of vehicle speed estimation, vehicle tracking and vehicle re-identification for highway traffic scenarios
- Part of team that represented Columbia University and secured 5th Place in Nvidia AI City Challenge 2018

COMPETITIONS

Third Prize: Columbia Engineering Projects Expo 2017 Sep 2017–Dec 2017

- Fused the artistic style of an artwork with content of an image using CNNs by maximizing the correlation in the generated image and feature map of VGG-19 layers. Improvised on it by object segmentation and videos

Runners Up: Columbia Data Science Society (CDSS) Hackathon 2017 (42 teams) Sep 2017

- Collaborated in a team of 4 to analyze Enron email corpus to find fraudulent behavior
- Used two-pronged approach of Network Analysis and NLP (Semantic Analysis + Topic Modelling)

Silver Medal: TFI Restaurant Revenue Prediction (Kaggle Challenge) Feb 2015

- Top 4% on private leaderboard for predicting restaurant sales and key drivers of the investments for a new restaurant location