

# VINAY KALE

vinay.kale@columbia.edu | vinaykale64.github.io

## EDUCATION

---

Columbia University New York, NY  
M.S. in Data Science, GPA: 3.82/4 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: Dask, Tensorflow, PyTorch, Keras, Git, Tableau, RShiny, Amazon EC2, Jenkins, MS Office

## PROFESSIONAL EXPERIENCE

---

CAPITAL ONE New York, USA  
Senior Data Scientist (CARD.AI Team) February 2019–Present  
• Part of a small team that works on automation of the account valuation and risk assessment framework that ensures the raw data gathered from financial data pipelines are redefined as end-to-end data science models that are ready to scale as needed  
• Work involves creation and maintenance of an in-house python environment via adding custom ML models, engaging in code reviews, monitoring source codebase via testing frameworks and keeping up the documentation.

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

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.  
• 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

## PUBLICATION

---

CVPR Conference 2018: Traffic Surveillance Research Workshop Paper New York, NY  
Research under Prof. Zoran Kotic 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 5<sup>th</sup> Place in Nvidia AI City Challenge 2018

## OTHER

---

Teaching Assistantships at Columbia University  
• Course Assistant for *COMS 4996 Applied Deep Learning* under Prof. Joshua Gordon Aug – Dec 2018  
• Course Assistant for *COMS 4995 Applied Machine Learning* under Prof. Andreas Mueller Jan – May 2018

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 for a new restaurant location