Vinay Dinkar Kale

Email: vinaykale64@gmail.com
Mobile: +1-646-510-1024
Portfolio: vinaykale64.github.io
Github: github.com/vinaykale64

Education

- **Columbia University, New York** — *NYC, New York*
  M.S. Data Science — *2017 - 2018*
  Relevant Coursework: Probability Theory, Statistics, Machine Learning, NLP, Deep Learning, Algorithms, Data Storytelling

- **Indian Institute of Technology Madras (IIT Madras)** — *Chennai, India*
  B.Tech Mechanical Engineering, M.Tech Product Design — *2011 - 2016*
  Relevant Coursework: Computational Engineering, Time Series Analysis, Regression Models, Operations Managment

Skills

- **Languages** Python, R, SQL, Spark, Scala, Julia, C, C++, LaTeX
- **Frameworks** Scikit, Dask, Plotly, Dash, TensorFlow, Keras, PyTorch, RShiny, Django, Flask, Tableau, MS Office
- **Platforms** Docker, GIT, Airflow, MySQL, Linux, AWS, GCP, Microsoft Azure, Heroku, Jenkins, Travis

Experience

- **Capital One, New York**
  Senior Data Scientist — *2019 - Present*
  - Worked on building cloud-based end-to-end credit card underwriting ML model deployed via a dockerized Python API, affecting more than 100M Customers, with estimated incremental value of $35M per year
  - Advised on design refactor of internal python libraries with focus on design (reusability, architecture, code quality)
  - Led development of modernized model monitoring and validation framework, decreasing the time and effort required to complete compliance-required monitoring from a week to 8 hours
  - Coached peer data scientists on effective developer habits with focus on code quality, consistent user documentation, unit testing and git workflows

- **Spotify, New York**
  Data Science Intern — *Summer 2018*
  - Worked closely with cross-functional teams of analysts, user researchers, product owners and engineers to develop a fan-artist pair segmentation pipeline which quantified the affinity for 20 billion fan-artist pairs
  - With that, analyzed how a new video overlay feature affects a typical fan-artist journey and how different fan segments engage with it, by statistical analysis on data from AB tests
  - Other Projects: Optimised data pipelines through Google Bigquery, Redesigned artist tiering dashboard

- **ZS Associates, India**
  Data Scientist — *2016 - 2017*
  - Built a intelligent Medical API using patient-data vector embeddings (similar to Word2Vec) trained on sequential big medical data. This served as input to several patient-level classification models
  - Calculated network influence in social media space (Twitter) for particular industry using graphical algorithms like Pagerank, following in-depth social media content analysis
  - Created a marketing-mix based spend analyzer data pipeline which calculates impacts of different marketing campaigns

Research

CVPR Conference 2018: Traffic Surveillance Research Paper

- Novel method of using Mask-RCNN for object detection and localization, vehicle tracking and vehicle re-identification for highway traffic scenarios.

Projects

Euphony: Python Package on PyPI

- Get alerts while running time-consuming code. With a simple wrapper, it plays classical music while the code completes. It also lets user choose among artists. Future development includes ability t play external mp3 files.

Market-Monitor: Live web-app to monitor financial markets

- Study stock and options prices with latest news for any company. The app uses Yahoo Finance APIs for data, Dash for visualization and Heroku for app deployment. It comes with a developer guide for documentation.

Volunteer Experience

Organizing Committee, PyData Conference NYC — *Summer 2019*
Led team for reviewing proposals and spearheaded the Diversity Scholarship Initiative

Teaching Assistant, Columbia University — *2018*
Applied Machine Learning, Applied Deep Learning.