Vinay Dinkar Kale

Portfolio: vinaykale64.github.io
Github: github.com/vinaykale64

Email: vinaykale64@gmail.com
Mobile: +1-646-510-1024

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
- Led development of agile model monitoring and validation framework with domain experts, decreasing the time and effort required to complete compliance-required monitoring from a week to 8 hours.
- Advised on refactor of python libraries from partner teams with focus reusability, architecture and design.
- Coached junior data scientists on developer habits like code quality, testing, documentation and git workflows

Spotify, New York
Data Scientist — *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 for Pharmaceutical clients using patient-data vector embeddings (similar to Word2Vec) trained on sequential big medical data. This served as input to patient-level classification models.
- Led project in a client-facing role for marketing-mix based spend analyzer pipeline which calculates impacts of different marketing campaigns. Helped secure continued engagement for similar projects down the line.
- Spearheaded effort within the company for reproducibility by on-boarding all projects with git version control and setting precedent for consistent user and developer documentation.

RESEARCH

CVPR Conference 2018: Traffic Surveillance Research Paper
- Novel Mask-RCNN method for object detection and localization, vehicle tracking and vehicle re-identification.

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.