Vinay Dinkar Kale

Email: vinaykale64@gmail.com
Mobile: +1-646-510-1024
Portfolio: vinaykale64.github.io
linkedin.com/in/vinay-kale-71b7b48a

EDUCATION

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

Indian Institute of Technology Madras (IITM) — *Chennai, India*
B.Tech Mechanical Engineering, M.Tech Product Design — *Aug 2011 - May 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 — *Feb 2019 - Present*
- Worked on building multiple end-to-end credit card underwriting XGBoost ML models deployed via Python API, reaching 100M+ Customers, with estimated incremental value of $35M per year
- Architected and implemented agile monitoring and validation framework for valuation and behavioural models, decreasing the time required to complete compliance-monitoring from a week to 8 hours.
- Built python libraries for Data Sourcing and Cleaning with focus on re-usability, architecture and design. Went through multiple refactor cycles to keep it efficient and up-to date.

Spotify — *New York*
Data Science Intern — *Jun 2018 - Aug 2018*
- Working with cross-functional team of user researchers and product analysts, developed a fan-artist pair segmentation pipeline which quantifies 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.
- Assisted on building optimised data pipelines through Google Bigquery and redesigning artist tiering dashboard.

ZS Associates — *Pune, India*
Data Scientist — *June 2016 - July 2017*
- Developed marketing-mix modeling pipeline using multivariate regression for client in retail media sector. Increased Total ROI by 15 percent and secured engagements for 3 similar additional projects.
- Built a Next-gen Patient Clustering Engine for pharmaceutical clients using patient-data vector embeddings trained on sequential big medical data. This served as input to several patient-level classification models.
- Led initiative for reproducibility within the company by on-boarding projects with git version control and a fostering culture of consistent user documentation.

RESEARCH

CVPR Conference 2018: Traffic Surveillance Research Paper
- We introduce novel usage of Mask-RCNN algorithm for vehicle detection, speed estimation and vehicle tracking for video footage of cars in highway. With it, we secured 5th place in Nvidia AI City Challenge 2018.

PROJECTS

Euphony: Python Package on PyPI
- Built a utility which lets users track code completion. With a simple wrapper, it plays classical music while the code is executed. Users can choose among range of artists.

Market-Monitor: Live web-app to monitor financial markets
- Developed an app which lets user study stock and options prices with latest news for any company, all in one place. It uses Yahoo Finance APIs for data, Dash library for visualization and Heroku for app deployment.

VOLUNTEER EXPERIENCE

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

Teaching Assistant, Columbia University — *Jan-Apr/Aug-Dec 2018*
COMS 4995 Applied Machine Learning, COMS 4996 Applied Deep Learning.