

Vinay Dinkar Kale

Portfolio: vinaykale64.github.io

Github: github.com/vinaykale64

Email: vinaykale64@gmail.com

Mobile: +1-646-510-10424

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
Teaching Assistant For: Applied Machine Learning, Applied Deep Learning
- **Indian Institute of Technology Madras (IIT Madras)** Chennai, India
B.Tech Mechanical Engineering, M.Tech Product Design 2011 - 2016

SKILLS

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

EXPERIENCE

- **Capital One, New York** 2019 - Present
Senior Data Scientist
 - 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** Summer 2018
Data Science Intern
 - 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** 2016 - 2017
Data Scientist
 - 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** Used Mask-RCNN for object detection and localization, vehicle tracking and vehicle re-identification for highway traffic scenarios. Secured 5th Place in Nvidia AI City Challenge 2018.

PROJECTS

Euphony

Python Package on PyPI

- The package alerts you when you finish running time-consuming code. It plays classical music while the code runs and stops when the code is done. Current version lets user choose artist optionally and future development includes ingesting an external url to any audio (.wav, .mp3) file.

Market-Monitor

Live Web-app to monitor financial markets

- The app lets user study stock and options prices on a chart across different time periods. It compiles latest news on the company for the user to stay up to date. The data is sourced through the yahoo finance and finviz APIs in backend, Dash library for front-end and Heroku platform for app deployment. Future development includes native options price calculator.

VOLUNTEER EXPERIENCE

- **PyData Conference 2019, New York** Summer 2019
As part of organizing committee, reviewed proposals and led the effort for Diversity Scholarship.