Vinay Dinkar Kale

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

Docker, GIT, Airflow, MySQL, Linux, AWS, GCP, Microsoft Azure, Heroku, Jenkins, Travis • Platforms

EXPERIENCE

Capital One, New York

Senior Data Scientist 2019 - Present

• Worked on building end-to-end credit card underwriting XGBoost ML model deployed via Python API, reaching 100M+ Customers, with estimated incremental value of \$35M per year

- Led development of agile Model Monitoring and Validation framework for valuation and behavioural models, decreasing the time required to complete compliance-monitoring from a week to 8 hours.
- Maintained python libraries for Data Cleaning and Feature Transformation with focus on reusability, architecture and design. Led the multiple refactors of the libraries to keep it efficient and scalable.

Spotify, New York

Data Scientist Summer 2018

- o 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
- o Other Projects: Optimised data pipelines through Google Bigquery, Redesigned artist tiering dashboard

ZS Associates, India

Data Scientist 2016 - 2017

- Led project in a client-facing role for marketing-mix pipeline which calculates impacts of different marketing campaigns. Helped secure continued engagement for 3 such projects down the line.
- o 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 the effort within the company for reproducibility by on-boarding 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 to 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