Vinay Dinkar Kale

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

Email: vinay.kale@columbia.edu
Mobile: +1-646-510-10424

EDUCATION

- **Columbia University in the City of New York**   NYC, New York
  M.S. Data Science   *Sep 2017 - Dec 2018*
  Courses: *Probability Theory, Statistical Inference and Modeling, Machine Learning, NLP, Deep Learning, Algorithms, Data Storytelling*

- **Indian Institute of Technology Madras**   Chennai, India
  B.Tech Mechanical Engineering M.Tech Product Design   *Aug 2011 - May 2016*

SKILLS

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

EXPERIENCE

- **Capital One**   NYC, USA
  Senior Data Scientist   *Feb 2019 - Present*
  - **Building ML Models**: Worked on building end-to-end ML models for account valuation deployed via REST API
  - **Design Refactor**: Led refactor of internal libraries to increase scalability and provide consistent UI
  - **Visualization Framework**: Created a model monitoring and validation framework template with Cookiecutter for adoption among different DS teams

- **Spotify**   NYC, USA
  Data Scientist Intern   *June 2018 - Aug 2018*
  - **Segmentation Pipeline**: Developed a fan-artist-pair segmentation pipeline which quantified affinity of a consumer towards an artist spanning over 16B such pairs
  - **AB Testing**: Analyzed how a new stories feature in app affects a typical fan-artist journey by leveraging AB Tests and using statistical significance analysis

- **ZS Associates**   Pune, India
  Data Scientist   *July 2016 - July 2017*
  - **Medical Clustering API**: Built a novice AI Medical API using patient-data vector embeddings (similar to Word2Vec) trained on sequential big medical data. This served as input to patient-level classification models.
  - **Mixed models**: Created a marketing-mix based spend analyzer data pipeline which calculates impacts of different marketing campaigns.

PUBLICATIONS

- **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 (April '18)

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 injecting an external url to any music 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 yfinance and finviz APIs in backend with the app deplyed on Heroku platform.

TEACHING AND VOLUNTEER EXPERIENCE

- **PyData Conference 2019**   NYC, USA
  [Organizing committee] Reviewed proposals and led the effort for Diversity Scholarship.   *July 2019 - Sep 2019*
- **COMS 4995 Applied Machine Learning**   NYC, USA
  [Teaching Assistant] Guided students and helped with evaluations.   *Spring 2018*
- **COMS 4996 Applied Deep Learning**   NYC, USA
  [Teaching Assistant] Guided students and helped with evaluations.   *Fall 2018*