PARITOSH KULKARNI

(646) 321-2902 | pck2123@columbia.edu | https://github.com/paritoshk | San Francisco

EDUCATION

Columbia University, *New York, NY*
Master of Science in Chemical Engineering, *Concentration in Data and Computational Sciences (3.6/4.0)*
Thesis: Modeling nonlinear dynamics of nonequilibrium nanoparticle systems in electromagnetic fields
Courses: Hedge Funds, Capital Markets & Investments, Game Theory, Data Analysis, Advanced Mathematical Methods & Modeling, Statistical Computing, Design Justice, Natural Language Processing
Honors: Columbia Build Lab'21 Fellow, Summer Institute on Sustainability & Energy Fellow (University of Illinois)

Manipal University, *Manipal, India*
Bachelor of Technology in Chemical Engineering, *Minor in Business Management*

SKILLS

- *Programming Languages* - Python (Numpy, Scipy, Pandas, Seaborn, Matplotlib, Spacy, Gensim, Transformers, NLTK, Regex, Scikit-learn, Keras, TensorFlow, Networkx), R, SQL
- *Enterprise*- AWS (S3, Sagemaker), Linux, GCP, PySpark, Plotly, Dask, Selenium, Streamlit, Hugging Face
- *ML/DNN Methods* - Topic Modeling, Word Embedding, Language Model, CustomBERT, DNN - Classification

EXPERIENCE

Pivotal Life Sciences, *San Francisco, CA*   Feb 2022 - Present
Data Scientist
- Created automated pipeline to encode gene relationships between 7 neurogenerative diseases into user-friendly no-code web application titled DealFlow InSight
- Engineered and deployed applications to search and filter six databases from 18 stock tickers from portfolio under assets over 2.2B +
- Mapped over 200k genes with appropriate locations, converted and identified three online databases with help of biologists to connect quantitative decision metrics used by investors using transfer learning

SAFELab at Columbia School of Social Work, *New York, NY*   April 2021 - Aug 2022
Lead Data Scientist
- Supervising and training two new data scientists and drafting weekly tasks to tackle NLP-experiencer problem
- Managing the overall data infrastructure in AWS, algorithms, and drafting methodologies, monitoring code
- Assisting in publication of two scientific articles through actionable data visualizations and literature contribution
Data Scientist
- Reconstructed - 34 GB of unstructured HIPAA data of over 20 million clinical notes - working with team of five
- Lead data cleaning, data frame reconstruction, risk factor analysis for human-generated notes & audio
- Utilized supervised learning to label- & text mine -medical records to locate instances of child abuse & neglect
- Identified 151 minors with positively labels out of 33961 in 2018, 92 % of them present in state criminal records

Siro.ai, *New York, NY*   Jan 2022 - Feb 2022
Machine Learning Engineer
- Defined binary classification NLP problem to annotate sales conversation transcripts between Sale or No Sale by creating a word embedding model and tested TF-IDF, n-grams, bag of words, ZeroShot and hugging face
- Created Deep Learning-based audio classification model for identifying instances of a change in pitch and confidence of speaker, from diarization to map frequency/timestamp with NER

ToughLeaf, *New York, NY*   Sep 2021 - Jan 2022
Chief Data Engineer
- Owning, planning, and debugging the data pipelines, dynamic data models, and AWS infrastructure
- Troubleshooting the existing database by removing duplicate entries, implementing combination algorithms
- Developing new databases for locations beyond New York by utilizing existing algorithms
Lead Data Scientist
- Applied NLP to condense & convert 10138 technical public data entries into actionable human-friendly form
- Collaborated with cross-functional teams to draft RFP for Empire State - Minority-Owned Business initiative
- Developed data cleaning algorithms to format & restructure data for front-end display at toughleaf.com