MRUGANK MILIND AKARTE

New York City (917) 900-8603
https://www.linkedin.com/in/mrugank-akarte/ mrugank.akarte@columbia.edu

EDUCATION

Columbia University New York City, NY
Master of Science, Data Science Dec 2020
Relevant Courses: Machine Learning, Exploratory Data Analysis and Visualization, Probability Theory and Statistics, Statistical Inference, Algorithms.

Vishwakarma Institute of Technology Pune, India
Bachelor of Technology May 2018
Bachelor of Technology in Production Engineering, GPA: 9.44/10.

WORK EXPERIENCE

Macy's Technology New York City, USA
Machine Learning Engineer- Product Recommendation Feb 2021-Present

As a Machine Learning Engineer at Macy's, I developed and deployed the "Complete the Look" recommendation model using Graph Neural Networks (GNNs) to enhance cross-selling opportunities for furniture products on macys.com. This resulted in 3-4% increase across different metrics like revenue per visit, average order value, and units per transactions. Other projects include product categorization and user category price prediction.

Key Accomplishments:

- **Model Development:** Built a GNN-based model that processes large-scale customer data, historical purchase behavior, and product attributes to create a comprehensive graph representation.
- **Real-time Inference:** Integrated the GNN-based model into the platform's real-time inference system, providing personalized recommendations to customers in seconds.
- **Pipeline Standardization:** Standardized multiple pipelines according to defined coding standards, migrating most from on-prem to Vertex AI, improving efficiency and scalability.
- **Operationalization:** Trained and supported data scientists in deploying models and pipelines using Kubeflow on GKE, ensuring seamless integration with the production environment. Developed notifications utility for teams and email on completion of pipeline for increased operational efficiency.
- **Innovation:** Conducted proof-of-concepts for parallel computing using Ray and explored ChromaDB for keyword and semantic search, demonstrating effectiveness in processing large datasets and enhancing search capabilities. Explored GKE cluster provisioning, containerized application deployment, and real-time monitoring using Prometheus and Grafana to ensure efficient scalability and high availability of machine learning workloads.

Technical Skills: Python, PyTorch, SQL, Google Cloud, Kubernetes, TorchServe, Ray, Docker

Ralph Lauren New York City, USA
Data Scientist (Capstone Project) Sep 2020 – Dec 2020

- Developed a machine learning model for the Ralph Lauren business team to predict the propensity of return for items in the cart.
- Investigated historical shopping and clickstream data to identify customer shopping and return behavior.
- Visualized multiple product features and their effects on return behavior.
- Designed models using AWS Sagemaker, AWS Athena, Python.

Nokia Bell Labs New Jersey, USA
Data Science Intern Jun 2020 - Aug 2020

- Devised an unsupervised CNN-LSTM autoencoder for anomaly detection and root cause(s) in multivariate time series.
- Model input were images of correlation matrices of different window lengths (short-medium-long term data) to capture spatial and temporal data simultaneously.
- Created model can be used to visually identify possible features responsible for anomalous behavior.
- Received Bell Labs summer intern award for outstanding innovation from the president of Nokia Bell Labs.

Indian Institute of Technology, Bombay Mumbai, India
Research Assistant under the guidance of Prof. N Hemachandra Jul 2018 - Jun 2019

- Explored ML and reinforcement learning theory and different methods to solve imbalanced classification problem.
- Published a paper for the ORSI conference on minimizing maintenance cost for Scania trucks utilizing the public dataset available on the UCI repository and observed up to 3x cost reduction when proper class weights model is utilized.

Ellicium Solutions Pune, India
Data Science Intern Jan 2018 - May 2018

- Conducted data cleaning, manipulation, and model testing for a customer retention project in the Insurance domain.
- Analyzed imbalanced datasets with low churn records, many missing values for certain variables, etc.
- Deployed logistic regression, SVM, Random Forest, Neural Networks, KNN, and gradient boosting techniques to keep a false positive rate within the expected range required by the client.
- Devised a submodule using R to capture sensory data from machines for real-time analysis.

LANGUAGE AND IT SKILLS

- Languages: Proficient in Python, SQL, R; Familiar with C++, Java, HTML.
- Packages: Tensorflow, numpy, pandas, keras, scikit-learn, dplyr, shiny, matplotlib, ggplot2, seaborn, plotly.

DATA SCIENCE PROJECTS

Columbia University, Model Quantization using TensorflowLite Dec 2020 - Dec 2020

- Explored different quantization techniques post-training quantization, quantization aware training, and weights pruning. Observed up to 4x reduction in model size with minimal reduction in model performance.

Columbia University, Exploratory Data Analysis on NATO Airstrikes Kosovo War Dec 2019 - Dec 2019

- Created various visualizations using heat maps, time series plots, and mosaic plots to understand the effects of NATO airstrikes during war on people. Built D3 visualizations depicting killings and casualties.

Kaggle, Toxic Comments Classifier Mar 2018 - Mar 2018

- Developed a multi-headed model capable of detecting different types of toxicity in a sentence using RNNs. Achieved AUC- ROC of 0.978.