ARUNIT MAITY

526 W 123rd Street
New York, NY, 10027

+1 (917)-592-9695
am5689@columbia.edu

EDUCATION

Columbia University — New York City, US
M.S. in Data Science — Dec 2022

- Relevant course work: Algorithms, Applied Machine Learning, Computer Systems, Data Analysis/Visualization, Practical Deep Learning System Performance, Probability and Statistics, Reinforcement Learning, Statistical Inference/Modeling

Vellore Institute of Technology — Vellore, IN
B.Tech in Electronics and Communication Engineering — Jul 2021

- CGPA : 9.33/10
- Relevant course work: Deep Learning, Neural Networks and Fuzzy Logic, Programming in C and Python, Data Structures and Algorithms, Data Science for Engineers, Essentials of Machine Learning

RELEVANT EXPERIENCE

Columbia University — New York City, US
Data Science Intern — Nov 2021 - Present

- Collaborating with the Environmental Flow Physics Lab on a research project to potentially shift 80-90% of US energy supply to renewable energy sources within the next 5-10 years
- Compiled historical US population and energy supply/demand data through web scraping and data cleaning using Python, NumPy and Pandas

National University of Singapore — Singapore, SG
Academic Intern — Jun 2019

- Received training from esteemed NUS faculty on Data Analytics, ML algorithms, Text Mining, Natural Language Processing and several ConvNet architectures such as VGG16, Resnet, LeNet-5, AlexNet etc.
- Partnered with other interns to design a recommendation system suggesting songs to a user based on current emotional state using NLP by emotion inference from user's social media posts through sentiment analysis

Hewlett Packard Enterprise — Singapore, SG
Academic Intern — Jun 2019

- Examined real-world datasets using the Apache Hadoop Eco-System and learnt how to set up Ambari clusters. Gained practical knowledge about MapReduce, Kafka, Hive, Sqoop, Flume, Cloudera Manager, Kerberos, Knox etc.
- Performed operations on an open source dataset of reservoir water levels in India and derived inferences using Apache Hive on real-time company servers and presented the findings to an HPE professional

SKILLS

SQL, MySQL, MATLAB, R, ggplot2, tidyr, dplyr, tidyverse, plotly, Python, NumPy, Pandas, scikit-learn, NLTK, BeautifulSoup, TensorFlow, Keras, PyTorch, Matplotlib, Hadoop, Hive, Impala, Spark, AWS, GCP, SAS, RDBMS

PROJECTS

Used Cars Price Prediction, Columbia University

- Trained an XGBoost Regressor to estimate second-hand car prices and showcase the effect of attributes through model interpretation. The model achieved a test R^2 score of 0.8074 and was deemed deployable due to low model complexity

NYC Traffic Accidents Analysis, Columbia University

- Examined the NYPD Motor Vehicle Collisions dataset with a team of 3 by visualizing multiple plots and maps with R packages - ggplot2, tidyverse, dplyr, tigris, cartography and geojsonio
- Highlighted important statistics/insights for prevention of collisions and constructed an interactive choropleth map on HTML with D3.js to show borough-wise collision count for varied scenarios

Lung Segmentation from CXRs using a DCNN, VIT Vellore

- Led a team of 3 to develop a DCNN for segmentation of lung parenchyma from CXRs facilitating lung diagnostics
- Optimized the system achieving a DSC of 0.982 ± 0.018 and a JSC of 0.967 ± 0.015. The paper will be published in the Biomedical Signal Processing and Control Journal (March 2022)

DTMF Tone Detection using the KNN algorithm, VIT Vellore

- Led a team of 2 to employ Goertzel's algorithm and KNN to design a model identifying DTMF tones with an accuracy of 98.47%. The paper has been published in Multimedia Tools and Applications Journal (July 2021)