Surabhi Ranjan

929-527-0283 | surabhi.ranjan@nyu.edu | linkedin.com/surabhi-ranjan/ | github.com/surabhi-ranjan

Education

M.S. in Data Science Sep 2020 – May 2022
New York University *New York, NY*
Coursework: Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Causal Inference

B.Eng. in Electronics & Instrumentation, Minor in Philosophy & Politics Aug 2013 – May 2017
Birla Institute of Technology and Science, Pilani, India *CGPA - 8.25/10.0*
Coursework: Probability & Statistics, Linear Algebra, Calculus 1 & 2

Technical Skills

Languages and Tools: PySpark, Python, SQL (Postgres), R, SparkSQL, Git, AWS, Databricks, Agile
Libraries and Frameworks: Pandas, NumPy, Scipy, Scikit-learn, Matplotlib, Apache Spark, Spark MLlib

Professional Experience

Data Scientist, Research and Development Jun 2017 – Aug 2020
MiQ Digital *Bengaluru, India*

AI-based Trading Assistant | *Reinforcement Learning, Multi-Armed Bandits*
- Contributed to key ML features in MiQ's B2B AI-based trading assistant that created advertisement-strategy recommendations and saved 10 person-hours of traders weekly.
- Optimized strategy recommendations through dynamic feature selection and ranked features of high dimensional sparse datasets using statistical techniques like Mutual Information, Chi-Squared.
- Automated the reallocation of daily budgets across advertising strategies using MAB-based reinforcement learning.
- Implemented a proof-of-concept using a LightGBM regression model to identify the optimal cost-per-advertisement value assigned to each strategy, reduced spend for a test campaign by 2%.

Hyperlocal Analytics Tool | *Python, PySpark, AWS, PostgreSQL, Redshift*
- Conceptualized and deployed, along with two fellow data scientists, a location-based advertising insights product to analyze user visitation and behaviour patterns; served 10M+ real-time in-app advertisements daily.
- Developed a geocoding algorithm that mapped client stores to potential target customers with a 70-meter accuracy.
- Led the development of production-level data pipelines using big data tools to support mapping of 50TB data to corresponding locations (>100k) and reduced data processing time from 3 hours to under 3 minutes.
- Optimized data storage of insight reports using AWS S3 and Redshift, and set up the management and monitoring of pipelines with AWS tools like EMR, Lambda, Cloudwatch, and SES.

Time-series Forecasting for Footfall Prediction | *ARIMA, PySpark, Tableau*
- Measured impact of online advertising on offline sales by observing offline visitation trends through a univariate, multi-step, ARIMA-based time-series forecasting tool.
- Leveraged Apache Spark to scale the aforementioned solution and run forecasting models in parallel for over 2 million locations spread across the US, reducing computing time by 20%.
- Created an interactive Tableau dashboard to increase interpretability of model output; used by 20+ analysts.

Projects

NYC Restaurants Health Inspection Violation Prediction | *NYU* Sep 2020 – Present
- Classified a restaurant into different severity levels of health violations. Implemented sentiment analysis and categorization of Yelp reviews using Natural Language Processing.

Fellow, Policy in Action | *Young Leaders for Active Citizenship* Jan 2019 – Mar 2019
- Co-authored a policy brief on the limitations of healthcare schemes through the lens of out-of-pocket expenditure and accessibility. Presented results and recommendations focused on optimized allocation and distribution of resources to the Government of New Delhi.

Founder, School Adoption Program, Nirmaan | *BITS Pilani, Goa* Aug 2013 – May 2016
- Connected CSR divisions of corporations with 5 municipal schools in Goa to bridge the infrastructural gap and lack of facilities faced by these schools which benefitted 3000+ students.
- Led a team of 20 volunteers to teach Mathematics and English to underprivileged children in Goa.