

Dharun Suryaa Nagarajan

San Jose, CA 95126 | (669) 243-7230 |

nagarajan.dh@northeastern.edu | [LinkedIn](#) | [GitHub](#)

Available May-Aug 2024

EDUCATION

Northeastern University, San Jose, CA

Sept 2023 - Current

Khoury College of Computer Science

Expected Graduation : May. 2025

Master of Science in Data Science

Related Course: Linear Algebra for Data Science, Collecting Storing and Retrieving Data, Supervised Machine Learning and Learning Theory, Data Processing and Data Management

Amrita School of Engineering, Coimbatore, India

July 2016 – May 2020

Bachelor of Technology in Civil Engineering

Related Courses: Database Design and Management, Object Oriented Design, Data Mining.

TECHNICAL KNOWLEDGE:

Programming Languages: Python, SQL, Shell Scripting, C++, R.

Statistics and ML: Hypothesis Testing, Time Series Analysis, ARIMA, Descriptive and inferential statistics, Regression, Classification, Clustering, Data Science Wrangling, Model Evaluation (ex: ROC, Accuracy, F-Beta, RMSE) , Feature Engineering, Ensemble Learning.

Tools: Microsoft PowerBI, Tableau, Snowflake, Data Studio, Jupyter Notebook, GitHub, Azure Databricks, AWS S3.

PROFESSIONAL EXPERIENCE:

Swiggy, Bengaluru, India

May 2022 – July 2023

Data Science Analyst

- Spearheaded the UE of Push Notifications by improving CTR and Uninstall rate by revitalization Python-based recommendation model, generating an additional 40k+ clicks per week and 16% reduction in Uninstall rate. Research involved feedback learning, PN-entity diversity and performance analysis and feature engineering(RFE, feature importance) Tools: Azure Databricks, AWS S3 and PySpark.
- Played an integral role in eliminating referral fraud which led to reduction in CTC of INR 63k per week. A rule-based model was built using 9 fraud features and their effects, determined using Log regression and ANOVA (data centralization, cleaning, EDA in pandas, Sklearn), on 20M+ referral records.
- Drove 5% enhancement in NU (New User) Activation per day by elevating App UI and sign-up flow. In charge of UI A/B Testing (experimental design and result analysis) on cohort of 250,000 NU traffic per day. Refined k-means clustering model on NU cohort based on historical order characteristics to facilitate cohort-based UI design.

Tata Consultancy Services, Bengaluru, India

Oct 2020 – Apr 2022

Junior Data Science Analyst

- Elevated the precision of Google Cloud Platform monthly recurring revenue (MRR) prediction accuracy by 13%. Orchestrated using advanced ARIMA forecasting model on the Google proprietary Dremel clusters connecting the end output to Google Sheets using custom Python connectors.
- Designed end to end GCC pipeline for the migration of 25 SQL scripts from Dremel to Flex Warehouse (Google Proprietary tools) and reduced run-time from 1 hr. to under 10 mins using SQL optimization techniques (indexing, memory optimized tables, conditional querying, and sequential joining)

ACADEMIC PROJECTS:

British Airways(BA) Customer Experience(CX) | Beautiful Soup, Requests Package, SciPy

- Scraped 3k+ user reviews on Sky Track site and applied schema standardization, format detox and EDA.
- Devised a metric wireframe and crafted a [PowerBI report](#) for stakeholders highlighting the insights from user reviews.

Extension Gmail Classification: | Rest API, TypeScript, XGBoost, Ensemble Learning

- Used Typescript and google client API to create an Extension that reads all the mails on a user's Gmail page.
- Used an ensemble model of Logistic Regression, Gaussian Naïve Bayes, and Random Forest to classify emails with a 91% accuracy, 0.85 F-Beta Score and a 0.868 as the ROC.