APARAJITA KAR

www.linkedin.com/in/aparajita-kar/
https://github.com/aparajitakar

Email : aparajitakar@icloud.com
Mobile : +1-651-269-4731
Address : San Francisco, CA

Education

- **University of Minnesota** — Minnesota, Minnesota
  Master of Science in Industrial and Systems Engineering (Data Science Track); CGPA:3.93/4.0 — *September 2019 - May 2022*
  - **Relevant Coursework**: Machine Learning, Data Mining, Applied Regression Analysis, Database Systems, Stochastic Models, Time Series Analysis, Analytics & Data-Driven Decision Making, Engineering Optimization
- **Visvesvaraya National Institute of Technology, Nagpur** — Nagpur,Maharashtra
  Bachelor of Technology in Chemical Engineering; GPA: 8.93/10 — *July 2013 – May 2017*

Technical Skills

- **Languages**: C, C++, Python, R Programming, MATLAB, SQL
- **Libraries**: NumPy, pandas, SciPy, scikit-learn, ggplot2, Matplotlib, Seaborn, dplyr, ggplot, tidyverse, lubridate
- **Tools**: AWS, Git, AMPL, Gurobi, Minitab, Tableau, Power BI, MS Excel
- **ML & Statistics**: Predictive Modeling, Risks, Hypothesis Testing, Metrics

Experiences

- **E2E Excellence LLC** — Remote
  Data Scientist — *September 2022 - November 2023*
  - Employed data extraction, analysis, interpretation, and data cleansing techniques on historical financial data to deliver actionable insights that optimized decision-making for the supply chain team using Python, SQL
  - Developed visually compelling and informative dashboard analyses in Power BI, effectively presenting complex data findings in an easily digestible format for stakeholders.
  - Leveraged diverse data sources to extract procurement data, utilizing this data to construct a methodology to predict sourcing time. This model empowered the supply chain team to make more informed sourcing decisions, increasing operational efficiencies and cost savings.
  - Collaborated with project managers and design executives to offer data-driven recommendations, improving project outcomes and product design.
- **Daikin Applied Americas** — Minneapolis, Minnesota
  Data Scientist Intern — *September 2020-December 2020*
  - Development of a regression model for optimizing shipment costs, resulting in a 10% cost reduction
  - Developed a DBSCAN clustering algorithm to identify optimal delivery hubs in the US and Canada, improving logistics efficiency.
  - Tableau software for visualization and reporting to senior management
  - Created Python scripts for team collaboration and demo software to streamline the logistical shipment process.
- **Base Education Ltd** — Bangalore, Karnataka
  Lecturer — *June 2017 -July 2018*
  - Mentored and guided students, fostering effective communication, problem-solving, and decision-making skills.
- **Hindustan Corporation Ltd** — Guwahati, Assam
  Process Intern — *June 2015 -July 2015*
  - Quantitative Study & performance of boiler operations and improve the efficiency of Multi-effect evaporators in the recovery pulping process and the bleaching section with process control assessment. Mechanical and process calculation, along with cost estimation and Risk assessment

Academic Projects

- **PIMA Indians Diabetes Classification ML Algorithm**: Developed a classification model for the early detection of diabetes mellitus by analyzing diagnostic measurements of patients. This project identified critical factors, including glucose levels and the Diabetic Pedigree Function, contributing to the likelihood of diabetes. Assessed the classification models using metrics like accuracy, sensitivity, and AUROC
- **Netflix Movie Recommender System**: Developed personalized recommendation systems by constructing Content-Based and KNN models, incorporating dimensionality reduction and Decision Tree algorithms. Data was cleaned and preprocessed using Python, TF-IDF values were computed, and cosine similarity measurements were used to enhance the accuracy of personalized suggestions.
- **NFL Rushing Yards Classification Models**: o Worked on understanding the yardage gained during open-rushing plays through feature-engineered models, utilizing a substantial dataset of 0.6 million records and 49 features. Multiple classification models were assessed using hyperparameter tuning, including Naive Bayes, Random Forest, XGBOOST, SVC, and Dense Neural Network
- **Sales Prediction using Advanced Regression Techniques**: Conducted comprehensive exploratory data analysis (EDA), encompassing data integration, handling missing values, outliers, and correlation analyses. Utilized R for feature engineering and data visualization through ggplots. Employed supervised machine learning models to gain insights into product sales variations across different stores.