Sudarshan Paranjape

Boston, MA, 02215
857-200-4575 paranjape.su@northeastern.edu linkedin.com/in/sudarshanp4 github.com/sudarshan4120

Education

Northeastern University **Expected May 2025**
Masters of Science in Data Science *GPA: **3.8/4.0***
Relevant Courses: Supervised Machine Learning, Natural Language Processing

Mumbai University **May 2022**
Bachelor of Engineering in Electronics and Telecommunications *CGPA: **9.44/10***
Relevant Courses: Database Management, OOPM, Applied Mathematics

Technical Skills

Programming Languages: SQL, Python, R
Visualization / Developer tools: VS Code, Power BI, Tableau, Git
Libraries: Matplotlib, Seaborn,Scikit-Learn, NumPy, Pandas, TensorFlow, PyTorch, NLTK
DevOps: AWS cloud (EC2, S3, RDS, EFS), Snowflake
Databases: MySQL, Oracle DB, MongoDB
Data Science/Analytics Techniques: Data collection, Regression, classification, Probability, Product Analysis, Statistical Analysis, Data Modeling, Bayesian inference

Experience

TryCatch Group **Jan 2023 - Jul 2023**
Data Science Intern *Mumbai, India*

- Orchestrated comprehensive data analysis of a sales dataset comprising **5M+** records using Python, encompassing data quality assessment, ETL and multivariate analysis
- Composed an interactive **Power BI** & **Tableau** dashboard to provide stakeholders with business strategy & actionable insights to enhance growth and sales, leading to a **15%** increase in product sales conversion rates
- Incorporated Power BI to perform data cleansing and Exploratory Data Analysis, creating compelling visualizations that improved identifying trends & data accessibility and reduced report generation time by 30%, enhancing user experience

Pacific Coatings **Feb 2022 - Dec 2022**
Data Science Intern *Mumbai, India*

- Spearheaded inventory management, streamlining operations and reducing stockpiling time of **250+** products by **15** days, resulting in enhanced inventory efficiency
- Teamed up with business stakeholders and subject matter experts in the cement industry to fine-tune time series forecasting models, adapting to the unique industry challenges, leading to a **25%** rise in monthly revenue

Projects

QueryFi: Generating Structured Queries from Natural Language | *NLP, Pytorch* – Link **Jan 2024**

- Implemented Text2SQL by fine-tuning state-of-the-art transformer models, **T5**, **BART** & **NMT**, to adeptly transform natural language into precise SQL queries
- Enhanced the robustness and precision of SQL query generation by leveraging **WikiSQL** corpus, with over 80,000 hand-annotated natural language questions, SQL queries, and 24,241 HTML tables
- Achieved a notable training accuracy of **80%** and a test accuracy of **96%** showcasing successful model training and optimization strategies for Text2SQL

Fraud Detection in Financial Transactions | *Python, SQL* – Link **Oct 2023**

- Executed Credit Card fraud detection on a heterogeneous dataset comprising of **10,000+** transactions & conducted feature engineering, data cleaning using **SQL** Database & python libraries like Numpy, Pandas & Spacy
- Implemented Cuckoo filters with a random forest classifier, establishing a robust two-step mechanism for effective fraud detection in financial transactions with an accuracy of **86%**

Optical Character Recognition for Sanskrit Language | *Python* – Link **Jun 2022**

- Developed a Deep Learning project to enhance accessibility of Sanskrit manuscripts, using a robust Convolutional Neural Network (CNN) that processes 32-word inputs for effective character segmentation
- Constructed a dataset with **100,000+** images across 40 intricate Sanskrit character classes, achieving 98% training and 94% testing accuracy using sophisticated feature extraction & classification methods