Shashank Dongre

Boston, MA | (857)-999-7525 | dongre.s@northeastern.edu | LinkedIn | GitHub | Portfolio

EDUCATION:

- **Northeastern University**, Boston, MA | Master of Science in Information Systems **May 2024**
- **Pune University**, Pune, India | Master of Business Administration in Marketing **Mar 2018**
- **COEP Technological University**, Pune, India | Bachelor of Technology in Mechanical Engineering **May 2015**

TECHNICAL SKILLS:

- **Programming**: Python, R, SQL
- **Data Analysis Techniques**: Statistical Analysis, A/B Testing, Data Visualization, Predictive Modeling
- **Databases & Tools**: MySQL, PostgreSQL, MS SQL Server, MongoDB, Excel, Tableau, Power BI
- **Data Science Techniques:** Machine Learning, NLP, Bayesian Hierarchical Modeling, Markov Chain Monte Carlo Simulations, Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Keras, Matplotlib, Seaborn, Plotly, NLTK, Spacy, Graphviz, Dplyr, Ggplot2
- **Version Control and Project Management**: Git, JIRA, Agile
- **Cloud Technologies**: Azure (Blob Storage, Cosmos DB, Data Factory, Data Studio, Synapse), Amazon Web Services (S3, Athena)

PROFESSIONAL EXPERIENCE:

- **Accenture Pvt Limited, Pune, India** (*Data Analyst)* **Aug 2018-July 2021**
  - Integrated sales and web data from Google Analytics and other social media engagement metrics by creating STTMs (Source to Target Mappings) and Data Models using **ER Studio** leading to **22%** improved data segmentation accuracy.
  - Utilized **SQL** to conduct traffic source analysis for e-commerce and retail client websites, reducing wasted spend on low-performing channels and boosting CVR (Conversion Rate) by 8%.
  - Utilized probabilistic models (BG/NBD, Gamma-Gamma) in **Python** aiding precise CLV (Customer Lifetime Value) calculations for retail and e-commerce clients, enhancing customer targeting through data-driven insights.
  - Created and visualized customer segments in **Tableau** using calculated fields, parameters, and sliders to analyze purchasing trends and assess the influence of external elements on marketing strategies.
  - Collaborated with cross-functional teams to identify their KPIs resulting in the development of 10+ customized dashboards for stakeholders using Tableau leading to a **15%** increase in data-driven decision-making.
  - Optimized SQL database performance, achieving a **20%** increase in query efficiency by enhancing indexes, streamlining complex queries, and implementing data normalization best practices, leading to improved system efficiency.
  - Collaborated with Business Analysts to translate and strategically align business requirements into technical specifications.
- **Persistent Systems Limited, Pune, India** (*Marketing Data Analyst*) **July 2017-July 2018**
  - Optimized SQL queries to increase efficiency by **30%** using recursive Common Table Expressions (CTEs), significantly reducing data access times, and accelerating decision-making processes.
  - Enhanced customer outreach by **40%** by employing data cleansing techniques in Excel and SQL, removing duplicates and correcting inaccuracies in e-commerce customer databases.
  - Utilized **Power BI**'s advanced features, including context filters and query parameters, to streamline report creation, resulting in a **100%** improvement in report generation speed.

ACADEMIC PROJECTS:

- **Precipitation and Snowfall Forecasting with Multi-Modal Architecture [*Link*]** **Apr 2024**
  - Developed an integration framework using **ConvLSTM** and **LSTM** to enhance precipitation forecasting from **~16K** satellite images and meteorological data for Lake Michigan.
  - Engineered a sliding window mechanism to bridge data gaps and tested various ML models, including Decision Trees and Random Forests, boosting prediction accuracy for lake effect snowfall.
  - Crafted an **encoder-decoder architecture** utilizing ConvLSTM2D and LSTM layers, regularized with dropout, culminating in a dense output layer for binary precipitation classification.
  - Achieved a F1-score of **0.71** and a recall of **0.85**, with the model accurately predicting rain with **94% probability**.
  - Demonstrated model's efficacy with significant gains beyond the **32nd epoch**, highlighting advanced skills in sequential data and multi-modal analysis.
- **End-to-End Data Solution for Instacart Product Recommendations on Azure [*Link*]** **Dec 2023**
  - Spearheaded the development, deployment, and maintenance of ETL pipelines in **Azure Data Factory**, processing over **1.6 million** records from relational databases and flat files to **Azure Cosmos DB**, significantly enhancing data ingestion efficiency, using **Azure Blob Storage** instance batches.
  - Architected and refined a scalable data solution in **Azure Cosmos DB** using **SQL API**, implementing indexing, partitioning, and query optimization strategies, which resulted in a **50%** improvement in query performance.
  - Engineered and deployed a hybrid recommendation engine in **Azure Synapse**, integrating **Azure Machine Learning** service to provide personalized product recommendations, enhancing customer experience.
  - Leveraged **Cosmos DB Gremlin API** to construct interactive dashboards and graph visualizations for near real-time analytics and insights into user-purchase patterns and behaviours, facilitating data-driven decision making.

Shashank Dongre

Boston, MA | (857)-999-7525 | dongre.s@northeastern.edu | LinkedIn | GitHub | Portfolio

- **Trend Prediction and Probabilistic Modeling of Live News Data using Neural Networks [*Link*]** **Apr 2023**
  - Designed an automated **data scraping** pipeline that retrieves real-time news data in natural language to implement rigorous data cleaning and preprocessing techniques, reducing data acquisition, and preprocessing time by **65%**.
  - Conducted exploratory data analysis (**EDA**) and sentiment analysis utilizing VADER and TextBlob libraries to evaluate news sentiment, enhancing data interpretation and preprocessing.
  - Employed **Latent Dirichlet Allocation** or LDA-based topic modeling, dimensionality reduction, and probabilistic modeling to uncover global themes, obtaining a **72%** prediction accuracy on unseen data categorization.
  - Developed and optimized a **Multi-Layer Perceptron (MLP) Classifier** for accurate trend prediction, refining the model through iterative tuning to reach an **87%** accuracy rate.
  - Developed interactive **Power BI** visualizations to analyse live news trends and sentiments across geographies.
- **Database Design and Analysis for Boston Real Estate Insights [*Link*]** **Dec 2022**
  - Automated data extraction from Zillow for comprehensive real estate analysis, covering thousands of properties listings.
  - Implemented data cleaning and preprocessing in **Pandas** to ensure accuracy and consistency.
  - Implemented data ingestion and management within **MySQL** Workbench, to Third Normal Form (3NF) for enhanced query performance, supporting 18+ analytical use cases with complex joins and subqueries.
  - Analyzed correlations and developed **Power BI** dashboards to enhance visual insights and support data-driven decisions.

CERTIFICATIONS:

- Supervised Machine Learning: Regression and Classification from Standford University and DeepLearning.AI, 2023 **[*Link*]**
- Advanced Learning Algorithms from Standford University and DeepLearning.AI, 2023 **[*Link*]**