**SKILLS:**

* **Programming:** Python, SQL, R
* **Data Modeling:** Statistical Analysis, A/B Testing, Predictive Modeling, Bayesian Hierarchical Modeling
* **Databases:** MySQL, PostgreSQL, MS SQL Server, MongoDB, Azure SQL Database, Azure Cosmos DB, Google BigQuery, Snowflake
* **Data Engineering & ETL:** Azure Data Factory, SSIS
* **ML/AI:** Scikit-learn, TensorFlow, Keras, NLP (NLTK, Spacy), Generative AI with Azure OpenAI (GPT-3.5/4), Azure AI Search
* **Data Visualization:** Power BI, Plotly, Matplotlib, Seaborn, Ggplot2
* **Cloud:** Azure (Azure Synapse Analytics, Azure Functions, Azure App Service, Azure Blob Storage), Google Cloud
* **Tools:** Git, JIRA, Agile, Figma, Balsamiq, Netlify, Google Analytics, Excel/VBA, REST APIs, CI/CD Pipelines

**PROFESSIONAL EXPERIENCE:**

* **Agile Datapro, Inc., Campbell, USA** (*Data Engineer*) **Aug 2024-Present**
* Engineered a scalable backend ecosystem with **Flask APIs** that route queries, integrate seamlessly with **OpenAI’s** **LLM** for dynamic recommendations and scoring, and execute **ETL** pipelines between source systems and **Azure Blob Storage**, and **Azure SQL**—ensuring real-time data integrity, availability, and **GenAI**-driven insights, improving response times by 50%.
* Migrated and optimized database infrastructure with role-based access control or **RBAC**, refining indexing and partitioning strategies to boost query speeds by 40%, reducing retrieval times by 30%, and optimizing stored procedures and queries for instant, analytics-ready data delivery.
* Implemented secure **ETL** pipelines for data transfer from multiple source systems to **Azure Blob Storage**, ensuring data integrity and achieving a 40% performance improvement in data workflows.
* Leveraged **Azure AI Search** with vector embeddings for precise document retrieval, centralized all indexed materials in Azure Blob Storage, and optimized server-side logic to dynamically generate files, resulting in a secure, scalable, and fully automated cloud-based deployment environment.
* Designed dynamic **React.js/Flask** user interfaces and dashboards, integrating real-time chat functionalities and data visualizations to improve user engagement by 25%.
* **Accenture Solutions Private Limited** (*Data Analyst)* **Aug 2018-July 2021**
* Automated **ETL** workflows with **SSIS** to integrate sales and web data from **Google Analytics** and social media into **BigQuery**, cutting loading time by 50% and boosting efficiency by 30%.
* Leveraged **SQL** in **SSMS** for initial data prep, then integrated and transformed data in **BigQuery** to analyze traffic sources—driving an 8% CVR improvement through data-driven insights and targeted marketing strategies.
* Designed and conducted **A/B tests** to evaluate personalized versus generic recommendation strategies, using Click-through Rate (CTR) and Conversion Rate (CVR) as key metrics and employed hypothesis testing for statistical significance.
* Utilized **Python's** SciPy and Statsmodels to perform hypothesis testing on user engagement data, applying BG/NBD and Gamma-Gamma models to calculate Customer Lifetime Value (CLV) to enhance data-driven decision-making.
* Collaborated with marketing to integrate personalized recommendations into digital ad campaigns, improving engagement by 15% and sales by 10% through strategic optimization based on user feedback and engagement metrics.
* Designed and implemented customer segmentation dashboards in **Power BI** utilizing **DAX** for calculated fields and Power Query parameters, analyzed purchasing trends to inform marketing strategies, and assessed external factors impacting consumer behavior.
* **Persistent Systems Limited** (*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 time, 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 to support targeted marketing efforts.
* Utilized **Power BI** to visualize key performance indicators such as Total Revenue and Average Order Value (AOV), resulting in a 100% improvement in report generation speed and drastically improving self-service analytics capabilities.

**ACADEMIC PROJECTS:**

* **Precipitation and Snowfall Forecasting with Multi-Modal Architecture [**[***Link***](https://github.com/ShashankDongre/Precipitation-and-Snowfall-Forecasting-with-Multi-Modal-Architecture)**]**
* Developed an integration framework using **ConvLSTM** and **LSTM** for enhanced precipitation forecasting, processing **~16K** satellite images and meteorological data for Lake Michigan. Engineered a sliding window mechanism and tested various ML models like Decision Trees and Random Forests to improve prediction accuracy.
* Crafted an **encoder-decoder architecture**, regularized with dropout, achieving an F1-score of **0.71** and a recall of **0.85**, with the probability predicting rain with **94%** accuracy.
* Demonstrated significant gains beyond the 32nd epoch, highlighting advanced skills in sequential data and multi-modal analysis, which improved lake-effect snowfall LES prediction accuracy.
* **End-to-End Data Solution for Instacart Product Recommendations on Azure [**[***Link***](https://github.com/ShashankDongre/Instacart-Product-Recommendation-System-and-Analysis)**]**
* Developed and maintained ETL pipelines in **Azure Data Factory**, processing over 1.6 million records to **Azure Cosmos DB**, enhancing data ingestion efficiency.
* Designed a scalable data solution in **Cosmos DB** using **SQL API**, improving query performance by 50% through indexing, partitioning, and optimization.
* Utilized **Cosmos DB Gremlin API** for interactive dashboards and visualizations, enabling near real-time analytics for user-purchase patterns and data-driven decision-making.
* **Trend Prediction and Probabilistic Modeling of Live News Data Using Neural Networks [**[***Link***](https://github.com/ShashankDongre/NLP-on-News-Articles)**]**
* Developed an automated web scraping pipeline for real-time news data acquisition and preprocessing, reducing data handling time by 65%. Conducted exploratory data analysis (EDA) and sentiment analysis using **VADER** and **TextBlob** for improved data interpretation.
* Applied **LDA-based** topic modeling and probabilistic methods for global theme discovery, achieving 72% prediction accuracy on unseen data.
* Optimized a **Multi-Layer Perceptron (MLP)** Classifier for trend prediction, refining the model to reach an 87% accuracy.
* Created **Power BI** visualizations to analyze live news trends and sentiments across geographies.

**EDUCATION:**

* **Northeastern University**, Boston, USA | Master of Science in Information Systems **[**[**Link**](https://graduate.northeastern.edu/program/master-of-science-in-information-systems-boston-5278/)**]**
* **Pune University**, Pune, India | Master of Business Administration in Marketing **[**[**Link**](https://pumba.in/default.htm)**]**
* **COEP Technological University**, Pune, India | Bachelor of Technology in Mechanical Engineering **[**[**Link**](https://www.coep.org.in/)**]**

**CERTIFICATIONS:**

* Supervised Machine Learning: Regression and Classification from Standford University and DeepLearning.AI, 2023 **[**[***Link***](https://coursera.org/share/5264c627b8610b04355cac80a24f7285)**]**
* Advanced Learning Algorithms from Standford University and DeepLearning.AI, 2023 **[**[***Link***](https://coursera.org/share/d060f2fbd14585bf363026bbd1d1a257)**]**