Ajay Deshpande

Boston, MA, USA | deshpande.aj@northeastern.edu | 7816559185
linkedin.com/in/deshpande-ajay/| github.com/Ajay-Deshpande

EDUCATION

Northeastern University, Master of Professional Studies, Applied Machine Intelligence (GPA: 4) Sep 2023 - May 2025
Courses: Data Management & Big Data, Decision Support & Business Intelligence, AI communication & Visualization

RNS Institute of Technology, Bachelor of Engineering in Computer Science Aug 2016 - Aug 2020
Courses: Machine Learning, Statistics & Probability, Linear Algebra, Artificial Intelligence

WORK EXPERIENCE

Data Scientist, Turing.com Palo Alto, CA| Nov 2021 - Aug 2023

Tech-Stack: Python, MySQL, Google BigQuery, Vertex AI, Mode (Data Visualization)

- Engineered features using Common Table Expressions (CTE) on **BigQuery**, registered them on **Vertex AI** Feature Store, reducing data serving latency for **ML models** by 14%.
- Developed and maintained consistent high-quality datasets using advanced SQL techniques, including **CTE, Window Functions**, and Stored Procedures on BigQuery, optimizing data analysis and reporting.
- Created **Jupyter Notebooks** on Mode to clean, transform, and aggregate data using python and pandas, ensuring data integrity and accuracy for **business intelligence reports**.
- Designed and developed multiple dashboards on Mode with **KPIs**, presenting insights and targeted lists of developers and consumers to focus team efforts on, increasing productivity by 20%

Data Engineer, Pecten Labs Bengaluru, India| May 2020 - Nov 2021

Tech-Stack: Python, MySQL, Tableau, Machine Learning, AWS Lambda, ETL pipelines

- Built a hierarchical Decision Tree Classifier chain using **Scikit-learn** to predict asset and product classes for financial entities, internalizing the process, saving $8.4M annually and improving the F1-score by 8%
- Conducted customer segmentation analysis using clustering techniques and detailed demographic profiling with Python and **Tableau**, enhancing **product recommendation** accuracy and user experience.
- Composed a **data pipeline** with Python scripts, Docker, Messaging Queues (Google PubSub), Cloud Run, Cloud Scheduler, and Build, processing 1500 news and Twitter **text articles** in 3 minutes
- Modeled **NLP models** using NLTK and TextBlob libraries on Vertex AI for performing sentiment analysis and topic categorization, and filtering irrelevant articles reducing memory utilization by 8% and processing time by 15%

PROJECTS

Expected Credit Loss (ECL) Modeling (Project Link)

Tech-Stack: Python, Machine Learning Pipeline, Logistic Regression, Scikit-Learn, Pandas

- Developed **Explainable ML models**, leveraging Bayesian probabilistic techniques to forecast Expected Credit Loss (CECL) by estimating Risk Transition, Exposure at Default, and Loss Given Default for loan portfolios, aiding financial risk management

Product Design A/B Testing (Project Link)

Tech-Stack: Python, Statistics, A/B testing, Power Analysis, Plotly

- Implemented A/B testing methodology to evaluate the impact of product changes on user engagement and revenue metrics. Utilized statistical analysis and **hypothesis testing** to make data-driven decisions and optimize product performance

Scalable-Time-Series-Forecasting (Project Link)

Tech-Stack: Python, Apache Spark, Time Series Forecasting, Meta Prophet, PyTorch (CNN, LSTM)

- Developed a scalable **time series forecasting** framework using **PySpark**, leveraging Facebook's **Prophet** library, CNN, LSTM, and SARIMA. Explored diverse methodologies to forecast time series data, providing insights into the performance and suitability of each approach for large-scale datasets

TECHNICAL SKILLS

Programming Languages: Python (Scikit-Learn, Pandas, NumPy, PyTorch, Keras, Flask, BeautifulSoup, Selenium)
Amazon Web Services: AWS Lambda, AWS SNS, SQS, S3, EC2, Redshift, DynamoDB, Glue
Google Cloud Platform: Cloud Run, Scheduler, Build, Functions, Storage, Vertex AI, Pub/Sub, Compute Engine
Databases: MySQL, MongoDB, PostgreSQL, BigQuery (CTE, Window Functions, Stored Procs)
Big Data Technologies: Apache PySpark, Apache Hive, Apache Kafka, Airflow
Tools: Excel, Google Sheets, Git/Github, Tableau, Power BI, Looker, Docker, REST API, Plotly