Ameya Deshmukh

Boston, MA | +1 (857)3967408 | deshmukh.amey@northeastern.edu | LinkedIn | GitHub |

Education

Master of Science in Data Analytics Engineering

Expected May 2025

Northeastern University, Boston, MA, GPA: 3.7/4.00

Coursework: Statistical Learning, Data Management, Computation and Visualization

Bachelor of Engineering in Electronics and Telecommunication

Aug 2018 - May 2022

Vishwakarma Institute of Information Technology, Pune, India, GPA: 3.76/4.00

Coursework: Machine Learning, Probability and Statistics, Data Structures and Algorithms

Technical Skills

Programming Languages: C++, Python(Pandas, Numpy, Scikit-learn), SQL, NoSQL, Unix

Databases: MS SQL Server, MySQL, Oracle DB, PostgreSQL

Visualization Tools: Tableau, Power BI, Visio, Microsoft Excel, Looker

Data Engineering: Database Design, Data Modeling, Data Warehousing, ETL, Big Data Technologies (Spark, Hadoop), Git **Data Science:** Classification, Regression, Neural Networks, Deep Learning, Predictive models, Quantitative models **Cloud and Infrastructure:** AWS (Redshift, S3, EC2, Lambda, Glue), GCP(Buckets, Datastore, Version control), Azure

Work Experience

Office Assistant, Northeastern University, Boston, MA | Python ,Excel(Pivot Tables),GCP,Flask

May 2024-Present

- Developed and deployed dynamic web applications using Flask, JavaScript enhancing website functionality and UI.
- Maintained Data ELT pipelines using Airflow scheduling for website's articles and news saving updating time up to 25%.
- Optimized querying using advanced SQL queries and improved workflow tracking efficiency by 20% using Git and Jira.
- Worked on GCP to maintain Datastores, Cloud buckets and version control of websites.

Data Analyst Intern, Newtuple, India | Python, LLM's, SQL, AWS EC2

Jan 2023 – May 2023

- Developed a Q&A Chatbot for an e-commerce platform using Python reducing business analysis time by 30%.
- Utilized LLMs ('Langchain') to convert Natural Language to SQL queries, simplifying database access for stakeholders.
- Automated ETL data pipelines with DBT by making Data models, achieving 50% time savings for pipelining.
- Cleaned and optimized data models ('inventory', 'customers', 'orders') in **PostgreSQL** for **AWS EC2** deployment.
- Applied time-series analysis and statistical modeling, leading to a 12% increase in sales and decision making.
- Created performance report for supporting product development using problem solving skills.

Machine Learning Intern, Autonise, India | Pytorch, Python, EC2, Flask

Jan 2021 - Mar 2021

- Trained a CNN model using PyTorch for a 'Quick Draw Recognizer'.
- Applied Standard Scaling and one-hot encoding in data cleaning and feature engineering process.
- Implemented image recognition functionality using OpenCV capable of identifying eight categories with 83% accuracy.
- Deployed the model on AWS's EC2 instance and used Flask for REST API connections to Angular JS.

Projects

Online News Popularity Classification | Link

- Identified and addressed highly correlated columns to improve data quality by performing data analysis.
- Applied cubed root transformation to handle right-skewed distributions performing data manipulation using ANNOVA.
- Reduced dataset dimensions from 27 to 15 features using PCA, capturing essential variance.
- Built a custom SVM classifier in Python using Sequential Minimal Optimization algorithm.

Seizure Prediction using EEG Signals | Link

- Developed a hierarchical CNN architecture with batch normalization, dropout layers for pattern recognition.
- Implemented preprocessing of data(label encoding, normalization) and designed a custom PyTorch class.
- Conducted rigorous model training with Adam optimizer and Cross-Entropy loss, achieving 74% accuracy.

Travel Data Management Application | Link

- Designed a relational database schema with interlinked tables implementing foreign key constraints.
- Orchestrated an intricate custom dataset leveraging Python for logically interlinking.
- Performed Advanced analytics SQL Queries like creating CTEs and subqueries for optimized and clean retrieval of desired results.
- Built a customer-facing web app using 'Streamlit' in python connected to MySQL database capturing customer data.

Recommendations System | Link

- Used Natural Language Processing (TF-IDF) to form vectors for K-Means and Hierarchical clustering.
- Cleaned and collected datasets from IMDB and Wikipedia.