Nagpritam Pradeep Naik

(+1) 469-363-4025 | prietam19@gmail.com | linkedin.com/in/nagpritam-naik/ | github.com/Nagpritam

EDUCATION

The University of Texas at Dallas **August 2023 - May 2025**
Master of Science, Information Technology and Management GPA: 4.0/4.0

- Coursework: Business Data Warehousing, Database Foundations for Business Analytics, Advanced Statistics for Data Science, Predictive Analytics for Data Science, Business Analytics with R, Big Data, AWS Cloud Architecture, Applied ML, NLP, System Analysis & Mgmt.

Visvesvaraya Technological University, India **August 2014 - July 2018**
Bachelor of Engineering, ECE GPA: 3.4/4.0

SKILLS

Programming Languages: Python, PySpark, Scala, SQL, Java
Data Tech Stack: Apache Spark, Kafka, Airflow, Hive, REST API, MySQL, Bigquery, Redshift, MongoDB, Pandas, Numpy, Snowflake, DevOps
Cloud Technologies*:* Azure(ADLS, ADF, Logic Apps, Synapse Analytics, Microsoft Fabric), AWS(S3, EMR, Glue, Redshift), Databricks, Power BI
Certifications: Microsoft Certified Data Engineer(AZ-900, DP-900, DP-203), AWS Certified Solutions Architect, Databricks Certified Developer Associate, Airflow Fundamentals

WORK EXPERIENCE

Volvo Group North America, Byhalia, MS, United States **May 2024 - August 2024**
Data Engineering Intern

- Constructed Data workflows to migrate data from WMS-Oracle Systems onto Azure Tables using Databricks Notebooks in the SML Team.
- Automate existing Python scripts to deliver webmail service using Logic Apps/Power Automate, with **33%** in operational efficiency
- Developed ADF pipelines, Databricks Notebooks, and Flask API calls for 3rd party application integration like SMTP, Microsoft 365 suite
- Used Web Scraping using Python scripts to extract attendance data and used Azure DevOps for Source Versioning and CI/CD releases

Walmart Global Tech India, Bangalore, India **May 2022 - June 2023**
Data Engineer III

- Contributed to the data lake for Data Governance reporting, modernizing 20+ complex Alteryx and 15+ DataStage ETL workflows to Spark-Scala code, and reduced the licensing costs by **$300,000** annually
- Developed streaming and batch ETL pipelines from diverse sources like CSV files, Kafka, and REST APIs, performed EDA on datasets
- Implemented workflows using Apache Airflow DAGs resulting in **27%** increased efficiency and change data capture in BigQuery tables
- Deployment, source code versioning and release into the test environment using Git, Concord/Jenkin software

Fractal Analytics, Bangalore, India **March 2021 - April 2022**
Data Engineer

- Led the development of a data lake on Azure Cloud, integrating supply chain, market, and finance data for an analytical tool providing prescriptive investment insights
- Designed and implemented a batch-processing ETL pipeline ingesting ~3 million records daily, achieving 97% accuracy and data quality.
- Developed and maintained ETL processes using Azure Data Factory, PySpark on Databricks, and medallion architecture, ensuring data harmonization, and deployment with CI/CD release pipeline using Azure DevOps

Tata Consultancy Services, Bangalore, India **October 2018 - March 2021**
System Engineer (Data)

- Constructed ETL pipelines in Azure Data Factory(ADF) and increased the efficiency to **~ 39%** compared to manual load, built data warehouse on Azure Synapse Analytics
- Optimized SQL stored procedures, queries, and user-defined functions, leading to **~20%** reduction in processing time by the API fetching

ACADEMIC PROJECTS

Growth Acceleration & Sales Optimization using Predictive ML **March 2024 - May 2024**

- Predictive ML project to accelerate growth and optimize the extensive offerings within the Meat Substitutes category for a popular US CPG brand using regression algorithms by analyzing different predictors contributing to its sales
- Data sourced from Circana's Point of Sales, Panel data, and US consumption trends, providing insights into consumption habits

Real-time streaming data using Microsoft Azure **April 2019 - July 2019**

- The project aimed at developing a PoC for producing IoT telemetry data using a Python console application
- Data in JSON format sent to the IOT hub aggregated on windows in Stream analytics, and sink being Cosmos DB with MongoDB API

Non-Destructive Automatic Fruit Detection/Separation Technology **Semester VII - VIII (Bachelors)**

- The project aimed at classifying fruits & vegetables into bad and good by sorting them based on sizes using Image processing algorithms
- Use of k-means clustering and edge detection technique, a hardware mechanism consisting of a conveyor belt for filtering out

ACHIEVEMENTS AND PARTICIPATION

- **Teaching Assistant**, for Spring Semester 2024 under Dr Sheen Levine, Visiting Associate Professor at the University of Texas, Dallas
- **Badgify** Award for Excellent Work at Walmart Global Tech India, Star Award at Tata Consultancy Services