

NITIN TEWARI

+19302159233 nitintewari39@gmail.com nitintewari39 git/nitintewari

Education

Indiana University Bloomington

Aug 2024 – May 2026

Master of Science in Data Science

Bloomington, USA

- Coursework: Machine Learning, Statistics, Deep Learning, NLP

SRM Institute of Science and Technology, Chennai

July 2018 – May 2022

Bachelors of Technology in Computer Science

Chennai, India

- Coursework: Statistical Analysis, Data Structures and Algorithms, Computer Vision

Experience

PwC

Sep 2022 – Sep 2023

Data Scientist

Bangalore, India

- Designed and owned production-grade **ETL pipelines** using **Python**, **SQL** and **Databricks**, ingesting ERP, SAP, and eCommerce data to support analytics and product reporting use cases
- Partnered with **product managers**, analysts, and business stakeholders to translate ambiguous business questions into well-modeled datasets, improving **forecast** accuracy by **15%**
- Orchestrated a hybrid recommendation system using **ALS** and **Apache Spark**, contributing to a **25%** sales lift via **A/B testing** and surfacing insights into user engagement and monetization
- Improved business **decision-making** by contributing to a **15%** improvement in forecast accuracy, by integrating **AWS S3** data sources with **BI** dashboards and collaborating directly with business stakeholders
- Leveraged **AI-assisted** coding tools (e.g., GitHub Copilot) to accelerate SQL query development, documentation generation, and data transformation logic
- Implemented **data quality checks**, validation rules, and monitoring, reducing downstream reporting errors and improving trust in core datasets
- **Python, SQL, Apache Spark, Databricks, ETL, Data Modeling, AWS S3, Data Quality**

PwC

March 2022 – Aug 2022

Associate Intern

Bangalore, India

- Performed **EDA**, **data cleaning**, and transformation on large operational datasets using Python and SQL
- Assisted in **data product support** and maintenance, performing **EDA** and validation checks to ensure accuracy across **ERP** and **Snowflake** systems
- Built **APIs** and tested integrations with **Postman**, reducing system errors by **18%**
- **Skills: SQL, Snowflake, EDA, Postman, API Testing**

Projects

Real-Time Data Engineering Pipeline for Reliability Analytics | *Kafka, AWS, Python, SQL*

May 2025

- Built a real-time **ETL pipeline** using Apache Kafka on **AWS EC2**, streaming and processing **10K+** events/hour with end-to-end latency
- Stored streaming data in **Amazon S3** and automated schema discovery and data lineage using **AWS Glue Crawlers & Data Catalog**, improving query efficiency by **30%**

MediBot - RAG | *LLM, NLP, AI, Langchain*

Sept 2024

- Utilized advanced **NLP** techniques with Python's spaCy and **NLTK** libraries to preprocess, tokenize, and analyze text data for accurate symptom recognition
- Deployed a GenAI-powered **RAG** application utilizing **FAISS** and **Langchain** to extract entities from electronic health records (**EHRs**), improving data retrieval speed by **40%** and reducing manual processing

Customer Segmentation using Clustering and PCA | *Python, Scikit-learn, Seaborn*

Dec 2023

- Applied **KMeans** and **DBSCAN** clustering on 10,000+ customer records, identifying 5 distinct behavior-based segments that improved targeting precision by 30%
- Reduced **feature dimensionality** by **60%** using **PCA** while preserving **95%** variance, and conducted comprehensive **statistical** data cleaning, boosting analytical accuracy

Technical Skills

Languages: Python, R, Pytorch, SQL, C++, C#, Java, PHP, HTML

Statistical Tools/Framework: TensorFlow, Scikit-Learn, Multivariate Regression, NumPy, Pandas, Matplotlib, SAP

Generative AI: GPT, Mistral, Llama, Langchain, Fine-tuning, Prompt Engineering, Claude

Big Data/Database: MySQL, Databricks, MongoDB, Google Suite, GCP, AWS, Hadoop, Azure

Visualization: Tableau, G Suite, Power BI, Microsoft Office Suite (Word, PowerPoint, Excel, Outlook), UI/UX