

# Artik Gupta

## Data Scientist

📞 8077780743 📩 artik.gupta.gupta@gmail.com 🤖 Github 💼 Linkedin 🌍 Gurugram

### SUMMARY

Data Scientist with **4+ years of experience** designing and deploying scalable machine learning models, data pipelines, and AI Agents using **Python, SQL, PySpark, and Agentic AI frameworks**. Built and optimized **classification and regression models**, and developed **agent-based AI systems** to automate data analysis and decision workflows. Hands-on experience in **feature engineering, model validation, hyperparameter tuning, and TB-scale data processing**. Strong background in **statistical analysis, SQL/PySpark-based big data extraction, and performance optimization**, with a proven track record of translating complex datasets into **production-ready, solutions**.

### KEY SKILLS

Data Analysis, Statistical Analysis, Agentic Ai, Exploratory Data Analysis (EDA), Data Modelling, Feature Engineering, Machine Learning Modeling, Deep Learning, Predictive Analytics, Data Wrangling, Model Evaluation & Validation, Business Problem Solving, Data Interpretation, Hypothesis Testing, Data Storytelling, KPI Development, Stakeholder Communication, Analytical Thinking, Performance Optimization, Root Cause Analysis

### TECHNICAL SKILLS

**Languages:** Python (scikit-learn, Pandas, NumPy, , tensorflow), SQL, MySQL

**Big Data & Processing:** PySpark, Spark SQL, DataBricks, Distributed Data Processing, AWS RedShift, Hive, AirFlow, Spark Streaming

**GenAi/AgenticAi:** OpenAi SDK, Fast-API, Asyncio, Logging, HuggingFace

**Machine Learning/Deep Learning:** Regression, Classification, Clustering, Time Series Forecasting

**Model Optimization:** Hyperparameter Tuning, Cross-Validation, Explaination

**Analytics:** Predictive Modelling, Quantitative Analysis, Statistics, Root Cause Analysis

### PROFESSIONAL EXPERIENCE

#### DATA SCIENTIST - ACCENTURE

Jun '24 - Present

- Optimized data interpretation & decision support by deploying **multi-step Agentic AI systems** using OpenAI GPT-5 & Python, achieving over \$50K in cost savings while seamlessly integrating outputs into production data pipelines & BI platforms
- Engineered scalable **PySpark data pipelines on Databricks**, processing **1M+ incremental records daily from GCP to Delta Lake**, enabling **reliable, high-performance analytics delivery**
- Implemented synthetic control testing using **Sparse Synthetic Control** and regression techniques, improving OpenAI GPT- and Python performance evaluation accuracy by 25%
- Performed **demographic, affinity, time-series, attachment rate, and take-rate analyses**, delivering actionable insights into customer purchasing behavior to optimize promotional strategies
- Enhanced store segmentation using **K-Means, DBSCAN, and hierarchical clustering**, improving targeting precision by **30%** and supporting **data-driven marketing decisions**

#### DATA SCIENTIST - EXL SERVICES

Dec '21 - May '24

- Orchestrated **SQL procedures with CTEs and UDFs automating reports**, cutting generation time 70% using Microsoft SQL Server.
- Built **BERT-based encoder model enabling one-to-one text mapping**, improving semantic consistency and downstream model performance significantly overall.
- Crafted a **price optimization model utilizing XGBoost & time-series forecasting** to bolster forecasting accuracy & planning reliability significantly

### PROJECTS

#### Stock Price Prediction using Sentiment Analysis

- Evaluated and compared the performance of BiLSTM and XGBoost models for predicting stock prices of ADANIGREEN, with a focus on assessing the impact of market sentiment features on overall model accuracy and performance.
- Used FinBERT for sentiment analysis complemented by feature engineering to create sentiment feature.

### EDUCATION

#### LJMU M.S. Data Science

May '23 - May '25

- Data Science

#### BITS PILANI, Pilani Campus - B.E. Civil Engineering

Aug '16 - Sep '20

- Secured **7.7 CGPA**