

SIDHESHWAR SHARMA

DATA SCIENCE ENGINEER

PUNE • +91-9970907353 • sidheshwarsharma77@gmail.com

[LinkedIn](#) • [GitHub](#) • [LeetCode](#) • [GeeksforGeeks](#)

PROFESSIONAL SUMMARY

Data Science Engineer skilled in **statistical modeling**, **causal inference**, and **ML pipelines** using Python SQL. Passionate about applying data to solve complex problems and drive business impact.

WORK EXPERIENCE

Machine Learning Intern

Defence Research and Development Organisation (DRDO) - ACEM, Nashik

June 2025 - Present

- Designed ML pipelines to predict material properties from experimental data.
- Applied feature engineering using physical equations and statistical methods.
- Achieved strong accuracy ($R^2 = 0.98$ standard, 0.99 with Repeated K-Fold Cross Validation).
- Contributed to an upcoming research paper with DRDO scientists.

TECHNICAL SKILLS

Languages: Python, SQL, C++

Machine Learning: Regression, Classification, Decision Trees, Random Forest, Gradient Boosting, Ensemble Methods

Statistical Methods: Hypothesis Testing, Causal Inference, Time Series Forecasting

Libraries & Tools: Scikit-learn, Pandas, NumPy, Matplotlib, TensorFlow, Keras, Git, GitHub

Gen AI: Retrieval-Augmented Generation, OpenAI API, Prompt Engineering

PROJECTS

Customer Churn Prediction

Python, Scikit-learn, SQL, Random Forest, Logistic Regression

- Built a machine learning model to predict customer churn using transaction and engagement data.
- Applied feature engineering (tenure, frequency, payment delays) to improve accuracy.
- Achieved **88% F1 score** and delivered actionable insights for customer retention.

Recommender System for Personalized Learning

Python, Surprise, Matrix Factorization, SQL

- Designed a collaborative filtering recommender system to suggest personalized learning content.
- Compared matrix factorization, KNN, and hybrid approaches; improved recommendation accuracy by 20%.
- Integrated with SQL backend for real-time querying and scalable deployment.

Automated Grammar Scoring Engine

Python, Whisper, spaCy, DistilBERT, Random Forest

- Developed an audio-based grammar scoring system (speech-to-text + grammar analysis).
- Extracted contextual embeddings with DistilBERT and applied Random Forest regression.
- Achieved $R^2 = 0.94$ alignment with expert scoring.

PDF Question Answering using RAG

Python, LangChain, LlamaIndex, OpenAI GPT-3.5, Streamlit

- Built a Retrieval-Augmented Generation (RAG) system that answers queries based on uploaded PDFs.
- Implemented semantic indexing, embeddings, and chunk-based parsing for information retrieval.
- Designed a user-friendly Streamlit interface, demonstrating deployment of ML pipelines into applications.

EDUCATION

Bachelor of Technology in Computer Science (Specialization: Data Science) Sep 2022 - Oct 2026

DY Patil International University, Pune

CGPA: 7.5

Senior Secondary (Class 12 - CBSE)

May 2022

Kendriya Vidyalaya, Nashik

Percentage: 81%

ADDITIONAL INFORMATION

Languages: English, Hindi, Marathi

Soft Skills: Communication, Problem-Solving, Adaptability, Team Collaboration

Problem Solving: Competitive Coding, DSA, Time & Space Optimization

Certifications: Full Stack Web Development (100xDevs by Harkirat Singh)