

PRANAV DOKE

 [linkedin.com/in/pranav-d](https://www.linkedin.com/in/pranav-d)  pranavdoke634@gmail.com  +91 93224 61884



EDUCATION

Pune Institute Of Computer Technology

Bachelor Of Computer Engineering

June 2026

CGPA: 8.7/10

SKILLS

Languages: C++, Python, SQL, Java

Core CS: Data Structures, Algorithms, Object-Oriented Programming, Operating Systems, DBMS

Frameworks: Flask, Streamlit, React

Databases: MariaDB, MongoDB

Data Engineering: ETL Pipelines, Data Modeling, Apache Airflow, dbt, Batch Processing

Tools: Git/GitHub, Docker, VS Code

Big Data: Apache Spark (Basic), Apache Hadoop (Conceptual)

PROJECTS

AI-Powered Collaborative Architecture Planner | *Next.js, TypeScript, Gemini, Liveblocks*

July 2026

- Built a full-stack agentic planning app where a Gemini-powered AI agent generates system architecture diagrams from natural-language prompts via 2 background AI pipelines.
- Integrated Liveblocks to enable real-time collaborative editing — synchronized node/edge state, live cursor positions, and user presence across all connected clients.
- Engineered a second AI pipeline converting diagrams into downloadable Markdown specs, with versioned storage via Prisma/PostgreSQL and Clerk-secured auth.

End-to-End Banking Data Pipeline | *Airflow, dbt, Snowflake, Python, SQL*

January 2026

- Designed and developed a scalable end-to-end ETL pipeline simulating banking transactions with automated ingestion, transformation, and loading.
- Orchestrated workflows using Apache Airflow, built modular dbt models with data quality validations, and wrote integration-level pipeline tests for transformation accuracy.
- Loaded and optimized analytical datasets in Snowflake using incremental loading to improve query performance and scalability.
- Implemented monitoring, retry mechanisms, and SLA tracking for reliable pipeline execution.

GenAI-Assisted Data Cleaning Pipeline | *Python, OpenAI API, Pandas, SQL*

March 2026

- Built an automated data cleaning workflow combining LLM-generated cleaning rules with Python preprocessing.
- Leveraged GenAI models to identify schema inconsistencies, missing values, formatting issues, and duplicate records.
- Applied structured JSON cleaning plans to standardize datasets and generate reproducible data quality reports.
- Designed normalized SQL schemas applying DBMS concepts, and stored validated datasets for downstream analytics.

EXPERIENCE

AI Evaluator & Annotator | *Scale AI / Outlier*

- Evaluated technical LLM outputs and performed data annotation under strict guideline rubrics.

In-House Internship | *PICT*
Workflow Designer & JSON Export Tool

March-May 2025

- Built a drag-and-drop workflow designer using React, TypeScript and ReactFlow
- Engineered a JSON serialization engine converting visual graphs into backend DSL format
- Implemented custom node/edge renderers with inline editing and conditional routing

CERTIFICATIONS

Complete Data Science, Machine Learning, DL, NLP Bootcamp | *Krish Naik*

Generative AI Virtual Internship | *Google*