Parth Sharma

8076759218 / parthsharma23212@gmail.com / https://www.linkedin.com/in/parth-sharma-08b1b424b/|https://github.com/ParthSharma272

Education

University School of Automation and Robotics, GGSIPU — New Delhi, India
Bachelor of Technology in Artificial Intelligence and Data science — *Nov. 2022 – May 2026(Ongoing)*
CGPA: 7.5 (Till 6th sem)

Modern Era Convent School — New Delhi, India
10+2 Sr. Secondary Education in Science — *April. 2008 – May 2021*
12th Percentage: 84.6

Experience

Machine Learning Intern — *Jul, 2025 – Aug, 2025*

Prodigal AI — *New Delhi, India*

- Automated Startup Evaluation System: Designed and led development of a pipeline that analyzes pitch decks, financials, and investor updates to extract and score KPIs, reducing manual bias and time.
- Tech Stack & Orchestration: Built with Apache Airflow for orchestration, Docker for workflow management, OCR + RAG-based AI pipeline for KPI extraction, and MLflow for tracking confidence scores.
- Scoring & Insights: Implemented a 33-KPI scoring model tailored to Indian SaaS benchmarks, with results delivered via a Streamlit dashboard combining quantitative scores and qualitative AI insights.
- Leadership & Learning: Grew as a team lead, integrating machine learning, data pipelines, and product thinking into an end-to-end industry-ready system.

Machine Learning Summer Intern — Jun 2024 – July 2024

University School of Automation and Robotics, GGSIPU — *New Delhi, India*

- Applied AI in Bioinformatics: Built machine learning models for drug discovery in Leishmaniasis, a neglected tropical disease, achieving 90% accuracy with Random Forest using RDKit fingerprints on a wide Excel Dataset of over 270k+ rows.
- Advanced Molecular Representations: Engineered compound features through molecular fingerprinting (RDKit, Morgan, MACCS, Atom Pair) to capture structural and activity-driving patterns.
- Model Development & Evaluation: Trained and compared 36 models (classical + ensemble methods), supported by feature importance analysis, PCA/t-SNE visualizations, and performance heatmaps.
- Impact on Drug Discovery: Showcased ML's ability to accelerate discovery timelines, lower costs, and provide interpretable insights into structure–activity relationships.

Projects

Startup Health Engine|Python, SQL, Apache Airflow, Docker, MLFlow, Streamlit

- Automated Startup Evaluation System: Built and led a system that automatically assesses startup health by extracting key metrics from pitch decks, financials, and investor updates.
- Technology: Used Apache Airflow for workflows, OCR + RAG for metric extraction, and MLflow for tracking accuracy.
- Scoring: Designed a 33-metric scoring model customized for Indian SaaS startups, turning raw financial, growth, and retention data into clear health scores.
- Insights: Created an interactive Streamlit dashboard showing real-time scores, confidence levels, and AI-generated insights to support faster, data-driven evaluations.

Book Recommendation System|Python, Jupyter lab, Transformers, Vector Databases, Langchain, Gradio

- AI Book Recommendation Engine: Built a system that recommends books based on meaning, not just keywords, using embeddings and vector search.
- Data Pipeline: Processed thousands of book descriptions from Kaggle—cleaning, handling missing data, and generating embeddings for similarity matching.
- Smarter Recommendations: Integrated LLM-based classification to auto-categorize books and added sentiment analysis to capture tone and refine suggestions.
- User Experience: Delivered an interactive Gradio app where users enter a title or description and instantly get personalized book recommendations.

Technical Skills

Languages: Python, SQL , JavaScript, HTML/CSS, R
Frameworks: Apache Airflow, Pytorch, TensorFlow, Django, XGBoost, Flask, FastAPI
Developer Tools: Git, Github, Excel, Docker, Streamlit, Google Cloud Platform, VS Code, PyCharm
Libraries: Pandas, NumPy, Sk-Learn, Matplotlib, Seaborn, NLTK, Shap