

ROSHAN KUMAR

Madhubani, Bihar, India | 7644045760

✉ rk1861303@gmail.com | [linked in.com/in/roshan-kumar-670529317](https://www.linkedin.com/in/roshan-kumar-670529317)

github.com/Roshan-pro

PROFESSIONAL SUMMARY

Analytical and results-driven Data Scientist with 2+ years of experience in Machine Learning, Deep Learning, and MLOps. Proficient in designing end-to-end ML pipelines using ZenML, TensorFlow, and CI/CD workflows. Adept at turning complex data into actionable insights through NLP, statistical modeling, and scalable automation. Proven ability to build intelligent systems and chatbot-based tools using FAISS, Streamlit, and Flask.

EDUCATION

Indian Institute of Technology (IIT) Patna – Patna, Bihar

2024 – 2027

B.Sc. in Computer Science & Data Analytics

Related Coursework: Machine Learning, Deep Learning, Data Structures, Computer Vision

Thesis: Resume Classification using NLP Techniques

SKILLS & TOOLS

- Programming: Python (Pandas, NumPy, Scikit-learn), SQL, MongoDB
- ML & DL: Regression, Classification, NLP, TensorFlow, PyTorch
- MLOps: ZenML, MLflow, Docker, Git, CI/CD
- Visualization: Power BI, Matplotlib, Seaborn
- Frameworks: Flask, Streamlit
- Soft Skills: Analytical Thinking, Communication, Collaboration
- Languages: English (Fluent), Hindi (Native)

PROJECTS & EXPERIENCE

AI Resume Analyzer

Flask, KNN, TF-IDF, Gemini, Docker

- Built an intelligent resume classification and feedback system using Gemini AI and TF-IDF.
- Implemented category prediction, chatbot assistant, job matching via API, and analytics dashboard.
- Achieved 85%+ accuracy on real-world resumes.

PDF Reader Chatbot

Streamlit, PyPDF2, FAISS, Sentence-Transformers

- Developed a web app allowing users to upload and query multiple PDFs in real time.
- Used semantic search to match user questions to content using FAISS and SBERT embeddings.
- Extracted and displayed both text and image answers from PDFs.

Sleep Disorder Prediction Pipeline

ZenML, MLflow, KNN

- Designed a fully modular ML pipeline using ZenML with integrated data preprocessing and model training.
- Used MLflow to track experiments and automated deployment with Docker + CI/CD.
- Achieved 86.3% model accuracy on lifestyle health dataset.

ACHIEVEMENTS

Winner :

- IIT Patna Hackathon (2025)
- Top Performer in Kaggle Playground Series (S5E6) – Multi-class classification challenge.

Finalist : ZenML MLOps Challenge (2024)

Language: Hindi, English