Tushar Kumar Gautam

+91 9927639137 | tushargautamiitr@gmail.com | linkedin.com/in/tushar | github.com/tushar

Summary

Electrical Engineering undergraduate at IIT Roorkee with expertise in Data Science, NLP, and MLOps. Proficient in Python and C++, with hands-on experience designing scalable full-stack AI applications and optimizing end-to-end ML pipelines. Skilled in deploying production-ready models using tools like Docker and MLflow to solve complex engineering challenges.

Education

Indian Institute of Technology, Roorkee Uttrakhand,India
Bachelor of Technology in Electrical Engineering *2022 – Present*

Experience

Data Scientist May 2025 – July 2025
PocketFM *Remote*

- Analyzed large-scale user behavior data to identify early churn indicators and enable targeted retention strategies.
- Segmented 2M+ users into behavior-based cohorts to identify high-risk groups and enhance personalization efforts.
- Evaluated A/B experiments by comparing session, impression, and click metrics to drive data-driven product decisions.
- Built an NLP pipeline for 50K+ user reviews to extract sentiment and uncover key issues for feature improvements.

Projects

Phishing Website Detection System | *Python, Docker, AWS, Github Actions, CI/CD* (GitHub)Jan 2026–Jan 2026

- Built a configuration-driven end-to-end MLOps pipeline for data ingestion, validation, training, evaluation, and deployment.
- Implemented automated schema validation, data drift detection, and reproducible preprocessing pipelines with artifact versioning.
- Integrated MLflow for experiment tracking, model registry, and controlled model promotion.
- Deployed containerized services on AWS (EC2, ECR, S3) with CI/CD automation using GitHub Actions.

End-to-End Quality Prediction MLOps | *Python, ML, Flask, MlFlow, Dagshub* (GitHub) Jan 2026 – Jan 2026

- Designed and implemented an end-to-end MLOps pipeline for wine quality prediction, covering data ingestion, validation, transformation, training, and evaluation.
- Implemented schema-based data validation using YAML to enforce feature consistency and ensure data quality before model training.
- Integrated MLflow for experiment tracking, logging parameters, metrics, and artifacts to enable reproducible model development.
- Built a Flask-based prediction API to serve trained models for real-time inference through a web interface.
- Containerized the application using Docker and automated CI/CD workflows with GitHub Actions for consistent and reliable deployments.

Streamlit-based RAG Configurator & Analyzer | *LangChain, FAISS, GroqAPI, Streamlit* (GitHub)Jan 2026 – Jan 2026

- Built a RAG benchmarking tool using LangChain to optimize chunking strategies for PDF data.
- Engineered a precision pipeline combining FAISS vector search and FlashRank reranking.
- Automated accuracy scoring using an "LLM-as-a-Judge" framework via the Groq API.
- Developed a Plotly dashboard to visualize performance metrics and latency trade-offs.

Technical Skills

Programming: Python, SQL, C++
ML & Deep Learning: PyTorch, TensorFlow, Scikit-learn
NLP & GenAI: Transformers, LangChain, HuggingFace, RAG
MLOps & CI/CD: Model Versioning, Experiment Tracking, Pipelines, Docker, GitHub Actions
Backend & Serving: FastAPI, Flask, Streamlit
Data & Tools: Pandas, NumPy, Matplotlib, Seaborn, Git, Jupyter, VS Code, DagsHub