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

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

- Built a modular pipeline using Pandas and Scikit-learn for automated data ingestion, transformation, training.
- Integrated MLflow and DagsHub to track ElasticNet experiments, log metrics (RMSE, R^2), and manage model versioning.
- Built a custom validation system enforcing schema constraints to ensure data integrity prior to training.

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.

Credit Card Default Prediction | *Python, numpy, pandas, matplotlib, seaborn* (GitHub) Sep 2024 – Sep 2024

- Developed an end-to-end credit card default prediction model with EDA, preprocessing, and feature engineering.
- Handled class imbalance using SMOTE and compared ML algorithms to identify the best model.
- Optimized the model using the F2-score, selecting Random Forest for high recall.
- Built a complete ML pipeline with threshold tuning and validation on unseen data.

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