AYUSH RAI

Data Scientist | ML Engineer | Backend Developer

🌐 Portfolio | Ayush-31r | ayush-rai-v1 | ✉ rai4ayush@gmail.com | 📞 +91 6388361301

EDUCATION

Indian Institute of Information Technology Manipur   Imphal, Manipur
Bachelor of Technology in **Electronics and Communication Engineering**   Nov 2022 – May 2026

EXPERIENCE

Software Developer (Data & Backend)   **Mar 2025 – Jul 2025**
Securify

- Designed a data-isolated, **multi-tenant system** to ensure **secure data handling** and access control for **6+** roles.
- Built automated **data processing pipelines** using **Celery**, **Flower**, and **Go**, improving the reliability of background **data management** and system tasks.
- **Analyzed and optimized database query performance** for **50+ REST APIs**, reducing query response times by **10%** and improving access to critical data.

PROJECTS

InsightOps – Support Ticket Risk Analysis System 🔗   **Dec 2025**

- Designed an end-to-end **ML pipeline** to predict **high-risk customer support tickets** using NLP features extracted from **120K+ historical tickets**.
- Trained and evaluated supervised models (**Logistic Regression, XGBoost**) achieving **0.86 ROC–AUC** and **81% F1-score** on a held-out validation set.
- Implemented a **retrieval-augmented explanation layer (RAG)** to surface similar past cases and policy context, reducing simulated manual triage time by **32%**.

Parkinson's Disease Detector 🔗 🔗   **Oct 2025**

- Built a **CNN-based model in PyTorch** to detect Parkinson's Disease from patient speech using the **MDVR-KCL dataset (2K+ samples)**.
- Processed raw audio into **spectrograms and MFCC features** with normalization and augmentation to enhance robustness on noisy biomedical data.
- Achieved **90.2% accuracy** and an **F1-score of 0.90**, deploying an interactive **Streamlit demo** showcasing real-time voice-based disease detection.

FlowCast – Urban Mobility Forecasting 🔗 🔗   **Aug 2025**

- Predicted **average inter-ward travel time and fare** in Bangalore using **Uber Movement data** and ward-level **GeoJSON** boundaries with spatial–temporal feature engineering.
- Engineered features such as **Haversine distance**, hour of day, and rush-hour indicators; visualized congestion trends via **Plotly choropleths**.
- Trained and deployed a **Random Forest model** achieving **MAE ≈ 485 s** and $\mathbf{R^2 = 0.67}$, with distance contributing **76%** to variance; hosted interactive predictions via **Streamlit**.

TECHNICAL SKILLS

Data Science & ML:	**Python**, **SQL**, **NumPy**, **scikit-learn**, **EDA**, **Feature Engineering**, **Model Evaluation**
NLP & GenAI:	**TF-IDF**, **Embeddings**, **NER**, **RAG**, **Vector Databases**, **LangChain**
Frameworks & APIs:	**PyTorch**, **Pandas**, **TensorFlow**, **Django**, **FastAPI**
Cloud & MLOps:	**Docker**, **Git**, **GitHub**, **GCP**, **AWS**
Other Skills:	**Go**, **C++**, **JavaScript**, **React**, **Tableau**

ACHIEVEMENTS

- Solved **300+ coding problems** across Codeforces, LeetCode, and CodeChef.
- Achieved **3-Star CodeChef rating (1733 peak)** with a **global contest rank of 164**.
- **Winner, CodeRush 2024** – secured **1st place** in Competitive Coding (Non-CS category).
- **Champion, 3 CTF competitions** hosted by the Data Security Club, outperforming **1000+ participants**.
- More **than 50 Badges** collected by leveraging GCP services in **the Google Cloud Skill Boost Program**.