

TUSHAR JAJU

+91-9927280188 | tusharbrisingr9802@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

Data Science student with production experience in legal-tech automation and ML optimization. Built compliance dashboard (480x faster assessment), deployed 7B LLMs with 95% transcription accuracy, and engineered analytics platform analyzing 2,500+ events. Delivers scalable solutions with quantified impact: Rs. 50K-2L savings, 80% time reduction.

PROJECTS

DPDP Act Compliance Dashboard – Legal-Tech Automation

Dec 2025

GitHub: github.com/Tushar-9802/DPDPA | **Tech:** Python, Streamlit, SQLite, Plotly, PDF Processing

- **Challenge:** Manual compliance takes 40+ hours; 1000+ startups facing May 2027 deadline with Rs. 10L-250 Cr penalty exposure
- **Database:** Normalized 9-table schema handling 100+ concurrent assessments; Third Schedule auto-detection (e-commerce 20M+, gaming 5M+)
- **Pipeline:** PDF extractor processing 200+ pages of DPDP Act/Rules; 95% accuracy classifying 50+ requirements
- **Performance:** 480x speedup (40 hours to 5 minutes); <100ms query response; generates Excel reports with penalty exposure
- **Impact:** Local-first deployment for data sovereignty; saves Rs. 50K-2L per assessment vs. legal consultants

AI Video Transcription & Summarization Pipeline

July 2025 – October 2025

Tech: PyTorch, Hugging Face Transformers, Whisper, Streamlit, CUDA

- **Challenge:** Manual transcription costs Rs. 200-500/hour; existing tools fail on domain terminology and multilingual content
- **Solution:** Deployed 7B-parameter LLMs (Llama 2, Mistral) with 4-bit quantization for consumer GPUs (RTX 3060); integrated Whisper ASR for 50+ languages
- **Optimization:** Flash Attention reduced memory by 40%; batch processing achieves 95% accuracy processing 2-hour videos in 8 minutes (15x real-time)
- **Scale:** Processed 200+ videos (400+ hours); automatic DOCX export with timestamps and speaker diarization
- **Results:** 80% time reduction (6 hours to 70 minutes per video); handles accented English, code-switched Hindi-English, technical jargon

Tech Layoffs Analytics Platform (2020–2025)

Jul – Aug 2025

Live: tushar9802-tech-layoffs-analysis.streamlit.app | **Tech:** Pandas, Plotly, Streamlit

- **Challenge:** Fragmented layoff data lacking quantitative volatility metrics for predicting industry turbulence
- **Solution:** Aggregated 2,500+ global tech layoff events; developed 3 proprietary metrics: Layoff Instability Score (rolling 6-month variance), Fragility Index (frequency-magnitude correlation), Bounceback Score (hiring velocity)
- **Performance:** Sub-second query response for 2,500+ records; real-time filtering across 500+ organizations
- **Insights:** Identified Q1 2023 as 38% volatility peak; gaming sector 2.4x industry fragility; 98% data accuracy vs. SEC filings

EXPERIENCE

Data Science Intern – Vinkura AI

Aug 2025 – Dec 2025

- Built predictive pipelines on 100K+ records with SMOTE + ensemble methods; improved F1-score from 0.72 to 0.83 (15% gain)
- Automated SQL ETL workflows reducing manual reporting by 20% (8 to 6.4 hours weekly); maintained ML pipelines with MLflow tracking

TECHNICAL SKILLS

- **Languages:** Python, C, C++, SQL
- **Data & ML:** Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow, CUDA, Quantization (4-bit/8-bit)
- **Databases:** SQLite, SQL (Schema Design, Query Optimization, Indexing, Normalization)
- **Visualization:** Plotly, Seaborn, Matplotlib, Streamlit (Production Dashboards)
- **Development:** Git, PDF Processing (PyPDF2, pdfplumber), Document Generation (openpyxl, python-docx)

EDUCATION

- **Bachelor of Technology**, Computer Science and Engineering ABES Engineering College, Ghaziabad (AKTU) CGPA: 7.00

Relevant Coursework: Data Science, Machine Learning, Database Management Systems, Software Engineering

ACHIEVEMENTS & CERTIFICATIONS

- 5-Star Python Certification – 3-Star Problem Solving • HackerRank
- Solved 300+ algorithmic challenges across LeetCode (150+), HackerRank (100+).
- Python, SQL, Problem Solving (Verified) • HackerRank – CISCO Python – IBM Data Analytics (9-course)