

ELVIA DEY

+46769558357 | elviadey7@gmail.com | linkedin.com/elvia-dey | github.com/elviadey07

EDUCATION

Uppsala University <i>Master of Science in Data Science</i> - Data Science Student representative - Teaching Assistant: Scientific Computing for Data Analysis	Uppsala, Sweden Aug. 2025 – Present Apr. 2026 – Present Aug. 2026 – Present
Ashoka University <i>B.Sc. Computer Science; Diploma in Advanced Research (Magna Cum Laude)</i> <i>Thesis:</i> Agent-Based Framework for Epidemic Simulation on Heavy-Tailed Network Structures - Teaching Assistant: Discrete Mathematics, Symbolic Logic and Applications	Sonepat, India Aug. 2021 – May 2025 Aug. 2024 – May 2025

EXPERIENCE

Data Science Intern <i>Virtusa</i> - Built a time-windowed SLA breach-prediction pipeline and autonomous agentic auto-remediation layer on 14.6K IT incident logs, deploying deterministic routing and audit logging. - Developed an 8-stage B2B Churn Propensity Engine paired with a Gemini RAG Copilot and retention simulator, preserving £250M in projected CLV.	Jun. 2026 – Aug. 2026 Remote
Inclusion Consultant & Data Analyst <i>Lodha Genius Programme</i> Analyzed quantitative program metrics and 80+ student interviews to deliver data-driven recommendations on STEM curriculum impact and inclusivity strategy.	Feb. 2024 – Jun. 2024 Sonepat, India
DE&I Consultant & Data Analyst <i>Credera</i> Examined the relationship between gender, education, and hiring outcomes in Colombia, and developed a targeted DE&I strategy to reduce biases in hiring practices.	Jun. 2023 – Aug. 2023 Remote

PROJECTS

SLA Restore-Breach Prediction & Agentic Auto-Remediation (Random Forest, XGBoost, Agentic AI) - Built point-in-time feature snapshots from 14.6K audit-log records across six elapsed-time windows, flagging breach risk 2.3× over baseline with PR-AUC and SHAP-based driver analysis. - Built a zero-touch LLM agent that autonomously selects and executes remediation actions per ticket.	Jul. 2026 – Aug. 2026
B2B Churn Propensity Engine (XGBoost, GenAI Copilot, Synthetic Data Generation) - Built an 8-stage churn pipeline on 7K+ synthetic B2B accounts, achieving 0.956 ROC-AUC and 0.80 F1 with XGBoost. - Designed a rule-based Next Best Offer engine, a retention simulation quantifying efficiency gains vs. blanket outreach, and a Gemini RAG copilot grounding natural-language queries in pipeline outputs.	Jun. 2026 – Jul. 2026
GitHub Star Prediction Pipeline (Scikit-learn, Docker) - Engineered 6 predictive features from 26 raw GitHub API fields to identify forks count as the top predictor. - Trained and tuned Ridge, Random Forest, and Gradient Boosting regressors, achieving 0.35 Test R^2 on a 1,000-repository dataset.	May 2026
Reddit “Tech Hype” Analysis (Apache Spark, HDFS) - Engineered a 4-node distributed cluster on OpenStack to process 19GB of raw JSON data; achieved an 85% storage footprint reduction and 5x I/O speedup by implementing Snappy-Parquet conversion. - Designed a keyword extraction engine to analyze technology trends across 3.8M records using PySpark.	Mar. 2026

TECHNICAL SKILLS

Languages: Python, SQL, R, C, Bash	ML/DL: Scikit-learn, PyTorch, TensorFlow, XGBoost
Data & Viz: Pandas, NumPy, Matplotlib, Seaborn	NLP: spaCy, BERT, Transformers, NLTK
Databases/Tools: PostgreSQL, Neo4j, Git, Docker, Parquet	Big Data/Cloud: PySpark, HDFS, Hadoop, OpenStack
GenAI & Agents: LLM tool-calling, agentic workflows, RAG, prompt engineering	APIs/Platforms: Groq, Hugging Face, Gemini API

LANGUAGES

Native: English, Bengali, Hindi	A2 Level: Swedish
--	--------------------------