

TechVista Solutions

Pvt. Ltd.

📍 Bangalore, Karnataka, India
🌐 techvista.ai

Summer Internship Report

June – July 2026

Machine Learning & Data Science Division

Priya Sharma ✉️ priya.sharma@iitd.ac.in 📞 +91 98765 43210 🔗 linkedin.com/in/priyasharma

B.Tech. Computer Science & Engineering | Indian Institute of Technology, Delhi | Roll No. 2023CS10245

About the Organisation

TechVista Solutions Pvt. Ltd. is a Bangalore-based technology firm specialising in enterprise AI, predictive analytics, and data engineering. Founded in 2015, the company serves over 40 clients across banking, insurance, retail, and healthcare in India, Southeast Asia, and Europe. Its flagship product, **VistaPredict**, powers real-time customer analytics pipelines for several Fortune 500 insurers. TechVista was recognised as a NASSCOM Emerge 50 company in 2024.

The Machine Learning & Data Science (MLDS) division, where this internship was carried out, focuses on applied research in churn modelling, recommendation systems, and NLP-driven document processing. The team comprises 15 data scientists and 6 ML engineers, led by Dr. Rajesh Krishnan (Head of MLDS).

Internship Objectives

- Understand the end-to-end lifecycle of an industry ML project — from business-problem framing to model deployment.
- Develop a production-grade **customer churn prediction** model using ensemble learning techniques on real-world telecom data.
- Gain hands-on experience with data preprocessing at scale, feature engineering, and hyperparameter optimisation.
- Collaborate within an agile team using Git, Jira, and daily stand-ups.
- Document findings in a technical report and present outcomes to the senior leadership team.

Project Description

Customer Churn Prediction using Ensemble Learning

The project addressed client retention for a major Indian telecom operator with over 12 million postpaid subscribers. Monthly churn rates of 3.2% were costing the client an estimated € 4.2M in annual recurring revenue. The goal was to build a classifier that identifies high-risk subscribers at least 14 days before a likely churn event, enabling proactive retention offers.

Dataset. The training corpus comprised 18 months of anonymised subscriber records (approx. 2.4 M rows × 47 features) including call-detail records, data-usage logs, billing history, complaint tickets, and demographic attributes.

Methodology.

- EDA & Cleaning.** Identified 11 % missing values in billing features; imputed via MICE. Detected and winsorised outliers in data-usage and call-duration columns (99th-percentile capping).
- Feature Engineering.** Created 14 derived features including usage-decline slope, payment-delay ratio, complaint-frequency index, and a network-switching affinity score.
- Modelling.** Benchmarked Logistic Regression, Random Forest, XGBoost, and LightGBM. Built a **stacked ensemble** with LightGBM + XGBoost base learners and a Logistic Regression meta-learner.
- Evaluation.** Achieved 0.91 AUC-ROC and 0.83 F1-score on a stratified 20 % holdout set. SHAP analysis revealed *complaint frequency* and *data-usage decline* as the strongest churn predictors.
- Deployment.** Packaged the inference pipeline as a FastAPI microservice; containerised with Docker and deployed on AWS SageMaker endpoints for batch scoring.

Weekly Activity Log

Wk	Dates	Activities	Status
1	2–6 Jun	Orientation, HR induction, toolchain setup (Jira, Git, AWS console), NDA signing, meeting mentor & team.	Completed
2	9–13 Jun	Literature review on churn modelling; collected and documented the telecom dataset; initial data-quality audit.	Completed
3	16–20 Jun	Exploratory data analysis: class-balance check, correlation heatmaps, missing-value profiling, outlier detection.	Completed
4	23–27 Jun	Feature engineering sprint: usage-decline slope, payment-delay ratio, complaint-frequency index; train/validation split.	Completed
5	30 Jun–4 Jul	Baseline models: Logistic Regression, Random Forest; hyperparameter search with Optuna; initial AUC ≈ 0.84.	Completed
6	7–11 Jul	Ensemble stacking (LightGBM + XGBoost); calibrated probabilities via isotonic regression; AUC improved to 0.91.	Completed
7	14–18 Jul	SHAP explainability analysis; prepared model card & fairness assessment; Docker container for inference API.	In Progress
8	21–25 Jul	Final presentation to leadership; handover documentation; exit interview; certificate distribution.	Planned

 All tasks tracked in Jira (Project: `CHURN-2026`); daily stand-ups at 10:00 IST.

Skills Acquired & Proficiency Matrix

Skill	Proficiency
Python & Scientific Stack	Expert
Scikit-learn & ML Pipelines	Advanced
Pandas / NumPy	Expert
SQL & Data Warehousing	Advanced
XGBoost / LightGBM	Advanced
Data Visualisation (Matplotlib, Seaborn, Plotly)	Advanced
Git & GitHub (CI/CD)	Advanced
Docker & Containerisation	Intermediate
AWS SageMaker & S3	Intermediate
Technical Writing (LaTeX)	Intermediate
Agile / Scrum Practices	Advanced

Key Learnings & Outcomes

- **Real-world data is messy.** Over 40 % of the internship was spent on cleaning, imputation, and sanity-checking before any modelling could begin — a sharp contrast to curated academic datasets.
- **Ensemble methods deliver.** Stacking LightGBM and XGBoost improved AUC by 0.07 over the best single model, confirming that diversity in base learners matters.
- **Explainability is non-negotiable.** Business stakeholders cared far more about *why* a subscriber was flagged than about raw AUC. SHAP waterfall plots became the primary communication tool.
- **Soft skills amplified impact.** Daily stand-ups, sprint retrospectives, and the final C-suite presentation sharpened both technical communication and cross-functional collaboration.

Mentor Evaluation

Dr. Rajesh Krishnan

Head of MLDS | TechVista Solutions Pvt. Ltd.

Technical Proficiency	4.5 / 5
Problem Solving	4.0 / 5
Communication & Reporting	4.5 / 5
Teamwork & Collaboration	5.0 / 5
Initiative & Ownership	4.0 / 5
Overall Rating	4.5 / 5

Priya demonstrated exceptional aptitude for applied machine learning. Her stacked ensemble model outperformed our internal baseline and has been selected for production rollout in Q3 2026. She integrated quickly into the team, communicated findings clearly, and showed strong ownership of her deliverables. I recommend her without reservation for any data-science or ML-engineering role.

Mentor's Signature

Date: 25 July 2026

Conclusion

This eight-week summer internship at TechVista Solutions provided an immersive introduction to industrial machine learning. Moving beyond textbook models, I grappled with real-world data quality issues, navigated stakeholder expectations, and delivered a churn-prediction system that is now on track for production deployment. The experience reinforced my interest in applied ML and gave me a portfolio of practical skills — from Optuna hyperparameter tuning to Docker-based deployment — that will serve as a strong foundation for my final-year capstone project and future career.

 This report was typeset in $\text{\LaTeX}$ and is available at <https://letx.app>.