Abstract
We present a lightweight approach to customer churn prediction that achieves robust cross-domain generalization using a minimal set of carefully engineered features and a cascaded teacher-student framework (logistic regression → random forest → RNN, distilled to GRU) with knowledge distillation. Our method demonstrates strong performance across diverse datasets (0.798 AUC, 88.6% recall), capturing 88.6% of churners while maintaining high class-specific precision (0.57 for churn, 0.74 overall). A key contribution is the application of knowledge distillation to tabular data, a relatively unexplored area, where we demonstrate that a GRU-based student network can match or exceed teacher performance while improving generalization to unseen domains. By leveraging entropy-based distillation and strategic multi-stage feature transformations, we achieve effective transfer across heterogeneous data distributions, offering a practical and efficient methodology for churn prediction in data-scarce or multi-domain environments.