DEVELOPING SOFTWARE COMPONENTS FOR A CHURN PREDICTION PIPELINE

Han Zhiqiang, Cherednichenko O.
National Technical University «Kharkiv Polytechnic Institute», Kharkiv

Customer churn prediction plays a vital role in the strategic decision-making of subscription-based businesses and service providers. Identifying potential churners early allows companies to take proactive steps to retain customers, optimize marketing efforts, and reduce revenue loss. However, building predictive models for churn requires a complex pipeline involving data collection, preprocessing, feature engineering, model training, evaluation, and interpretation - tasks that are often inaccessible to users without a technical or data science background.

While many existing platforms, such as Amazon SageMaker, DataRobot, and H2O AutoML, provide automated machine learning (AutoML) solutions for churn prediction, they typically offer limited transparency or flexibility. These tools often assume that users have at least some understandings of machine learning workflows, which creates a barrier for domain experts who want to customize models or experiment with different scenarios without diving into code.

This research addresses the challenge by focusing on the development of modular software components specifically designed for churn prediction tasks. The goal is to create a flexible pipeline architecture that allows users to build, modify, and execute predictive workflows using simple configurations, rather than programming. Each component of the pipeline - from data cleaning and segmentation to model training and churn risk scoring - is designed to work independently and be reusable across different contexts and data sources.

The core contribution of this work is the implementation of a customizable pipeline system tailored for churn prediction, with an emphasis on usability, clarity, and adaptability. The system supports integration with commonly used data sources (e.g., CRM systems, customer interaction logs) and provides built-in templates for typical churn analysis scenarios, such as predicting voluntary cancellations or usage-based churn.

An experimental evaluation was conducted to assess the usability of the pipeline for non-technical users. This involved user testing with business analysts and product managers, measuring factors such as time-to-deploy, configuration complexity, and the interpretability of model outputs. Resource usage and scalability were also analyzed under different deployment conditions.

Future work includes the development of a user interface that enables drag-and-drop configuration of pipeline components, as well as deployment as a cloud-based service to ensure accessibility and scalability. By lowering the technical barriers to churn prediction, this project aims to empower a broader range of professionals to leverage predictive analytics in their everyday decision-making.