

In this project, we tackled the challenge of predicting user adoption for a product, using a dataset rich in user engagement and characteristics. The primary task was to identify an "adopted user," defined as someone who logs into the product on three separate days within a seven-day period.

Data Wrangling and EDA:

We commenced with data wrangling and exploratory data analysis (EDA), focusing on cleaning the data, addressing missing values, and transforming features such as timestamps into more informative metrics like user lifespan. These steps laid the groundwork for a detailed analysis of user behavior.

Model Selection and Training:

Various machine learning models were explored, including Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machines, each chosen for their distinct advantages. Logistic Regression provided simplicity and interpretability, while Random Forest and Gradient Boosting were ideal for their robustness in complex data relationships. SVM was included for its efficacy in high-dimensional spaces.

Target Variable Class Imbalance and SMOTE:

A significant challenge was the class imbalance in our target variable (is_adopted_user: No= 7601, Yes= 1222). To address this, we would consider using Synthetic Minority Over-sampling Technique (SMOTE) or related methods in future work. These techniques generate synthetic samples for the minority class, aiming to balance the dataset and improve the model's performance on underrepresented classes.

Feature Importance Analysis:

We conducted a detailed feature importance analysis, particularly for Random Forest and Gradient Boosting models. This analysis revealed that user engagement metrics, such as total logins and user lifespan, played a crucial role in predicting user adoption. Understanding these key drivers is essential for refining the models and focusing on impactful features.

In conclusion, while the initial phase of the project laid a solid foundation, additional approaches like addressing class imbalance with SMOTE, and extended feature engineering would further enhance model accuracy and insights, leading to a more effective strategy for increasing user adoption and engagement.

