Final project: Churn Prediction in Mobile Games

SAPIENZA
UNIVERSITÀ DI ROMA

Dila Aslan - 2113310
Selin Erol - 2053197
Umut Altun - 2101934

Task and Motivations

Task Statement:

1. Define user churn in the context of mobile games.
2. Explain the importance of predicting churn using machine learning.

Motivation

1. Discuss the impact of user churn on revenue and the importance of retention strategies.
2. Highlight the benefits of machine learning over traditional statistical methods for churn prediction.

Related Works:

1. Churn Analysis Using Deep Convolutional Neural Networks and Autoencoders (https://arxiv.org/pdf/1604.05377.pdf)
2. On Analyzing Churn Prediction in Mobile Games (https://arxiv.org/pdf/2104.05554.pdf)
3. Predicting Player Churn of a Free-to-Play Mobile Video Game Using Supervised Machine Learning (https://www.mdpi.com/2076-3417/12/6/2795)

Models and Tools

1. We will use the Python as the primary programming language for data manipulation and model building.
2. We will make experimentation with common supervised ML algorithms such Logistic Regression, Random Forest, and Gradient Boosting Machines like LGBM and XGBoost.
3. We will try some different techniques such as undersampling to deal with imbalanced dataset.
4. For hyperparameter tuning process, we will use grid search strategy.
5. We will evaluate the model with common 5-fold CV strategy.
6. For evaluation metrics we will use recall, precision, and F1 score due to the imbalanced nature of the dataset.
7. We will use SHAP framework for feature importance.

Analysis

1. Anonymized churn data of a mobile games will be used for training and evaluation. These datasets have three tables: users, sessions and purchases. (https://www.kaggle.com/datasets/mobilegameguru/anonymzed-mobile-game-user-data)
2. We will try to use different classification metrics such as recall, precision, F1 score, ROC-AUC score.When we are using these metrics, we will also consider the business context for each of them.
3. We will discuss how the current best model's performance stacks up against findings in similar studies.
4. We will also discuss the model robustness by testing model across different subsets of the data.
5. We will use the SHAP feature importance framework to interpret our models. Because identification of most influential features can provide great insights about the churn problem.
6. Lastly, we will address the our limitations of the current model and dataset and we will outline potential future improvement.

Churn Prediction in Mobile Games