Aryng

KIVA CASE STUDY

HOW ARYNG **IDENTIFIED $500K+** IN POTENTIAL INCREMENTAL REVENUE **USING LTV**

More info
www.aryng.com

INTRODUCTION

A Forbes study reported that 9 out of 10 startups fail and discontinue after a certain period. Harvard business review stated that more than two-thirds of them never deliver a positive return to the investors. Failory.com suggests that 30-40% of investors lose their entire initial investment. But why?

Some researchers blame the insufficient funds while others blame the competitive landscape. Some put it on irregular marketing spending, while others criticize market knowledge.

The reason could be anything but the fact is unnerving! This is, In particular, a bigger problem than just a startup's loss. One start-up failure may mean a loss for one venture capitalist, but what does 90% start-up failure imply? It implies economic slowdown, declined GDP, and unemployed, de-motivated youth for a nation at large.

Quite alarming!! I know

On the other hand, there are these other startups that succeed and witnessed rapid growth every year. One cannot help but wonder what they do differently than their competitors to perform so well? What puts them on the growth trajectory immediately after they take off?

Let us see what research has to say.

THE KEY DIFFERENTIATOR

One key differentiator is the way they leverage the data.

Successful startups quantify their key financial parameters by leveraging data analytics to identify their critical growth drivers. Statistics play an essential role in such quantifications. For example, organizations like StarBucks take advantage of the LTV metric to analyze their customer persona and drive increased financial success year after year. Every startup can succeed if it can focus on a relevant financial metric like LTV and include it in its overall marketing strategy.

THE KEY DIFFERENTIATOR

LTV is short for Life Time Value of a customer. LTV can be defined as the predicted monetary value a customer would generate over a period of time (like two years or five years), based on early indicators of value. LTV helps businesses identify high-value customers early in their journey with the organization.

LTV can be used throughout the customer lifecycle to:

- Deepen engagement and drive value.
- Optimize acquisitions to maximize ROI.
- Help focus on high-value customers for incentivizing them.
- Tailor marketing spends for increased returns.
- Enable aggressive churn mitigation for high-value customers.

Let's review the above concept with a client case perspective.

CASE STUDY

ABOUT KIVA

Kiva is a micro-financing, fintech company that enables donors worldwide to support the causes they care about and make a real personal impact on the lives of underprivileged entrepreneurs and students. Since Kiva engages in allowing the donors to lend money, the donors become the users of the fintech company.

We at Aryng, got an opportunity to empower Kiva with actionable insights that can increase its revenue and rapidly accelerate its financial growth. We used our proprietary five-step Data-to-Decisions framework called BADIR to solve Kiva's business problem.

BADIR is an acronym where B stands for the business question, A for the Analysis plan, D for data, I for insights, and R for the recommendation.

THE KIVA LTV MODEL PROJECT

Step 1 in BADIR Framework: Business Question

The first step in the BADIR process is to ascertain the right business question through the who, when, where, what, why, and how to approach. The right business question leads us to the heart of the issue. It helps in identifying the absolute zone where the action is needed.

For this process, we consulted with the head of product @Kiva, and other stakeholders and asked several relevant questions to understand the reason behind this project. Some areas covered included impacted segments, actionability, stakeholder roles and responsibility, and potential reasons behind the opportunity/problem. Conclusively, we identified the problem zone and the real business question.

What was the real business question?

Build a 2-year predicted LTV model for all Kiva global lenders to understand drivers of LTV towards growth (all source and channels).
LTV, the target variable (what we are predicting) for a user @ Kiva has been defined as the amount donated by the user to Kiva during or outside the loan process.

Step 2 in BADIR Framework: Analysis Plan

After ascertaining the real business question, we moved on to building the analysis plan

The role of the analysis plan is to identify the most influential set of parameters to solve a given problem and design a plan of action that correctly aligns the proposed solution with the stakeholder's need.

At this stage, we involved Kiva's stakeholders in creating a plan of action to ensure we aligned with their business goals and avoided a possible future argument.

Analysis Goal

Stakeholders at Kiva wanted a model that could output a series of logic in identifying high-value customers.

Methodology

Another key component of the project is choosing the suitable methodology for solving the problem at hand. The methods adopted by analysts differ according to the business problem's nature and the business constraints.

For example, machine learning models often have higher prediction accuracy than statistical models. However, they have low explainability. Therefore, we decided to use statistical modeling techniques (for interpretability) for kiva and then extend them to machine learning models for better accuracy.

Hypotheses

One of the most critical parts of an analysis plan is hypotheses. A hypothesis is a potential reason that may explain the business problem. Hypotheses are important because:

- They allow us to focus on the most promising areas to solve the business problem. They also help to avoid wasting time in studying tangential areas.
- They allow stakeholders to engage in meaningful ways as their gut sense of solving the problem is incorporated in the analysis.

After talking with stakeholders, some hypotheses we arrived at included:

- Corporate lenders have a lower LTV
- Subscribers have higher LTV
- Users from the US tend to have a lower LTV

Time Window

This is another critical component of an analysis plan. For Kiva, some vital time window details included:

- Users will be observed for a window of three months starting from their first transaction.
- Based on that behavior, we will predict their 2-year LTV from the time of acquisition.
- The users acquired between 2-3 years ago will be considered an in-time sample.
- The users acquired during the next month post the in-time sample will be considered the out-of-time sample (validation).

The other components of the analysis plan included detailed data specification, sampling, and a project plan.

Project kickoff

Eventually, after working with the stakeholders and the internal Kiva team to align everybody on every aspect of the analysis plan, we got the buy-in. We kicked off the execution of the project.

Step 3 in BADIR framework: Data Collection

This stage involved pulling data for the said one-year time period and validating the same. The data validation process includes sanity checks apart from checking for the irregularities and accuracy of data. This is also called data cleansing.

We acquired the data for the said one-year time and cross-checked it for accuracy.

Step 4 in BADIR Framework: Derive Insights

a) Data Preparation

Deriving insights is the next step that starts with data prep to improve data quality. Data prep includes dropping missing values (eliminating smaller data sets to avoid biased parameters and estimates during analysis).

The data preparation process revealed that most variables had a very high percentage of missing data, including the target variable. The follow-up consultation with Kiva revealed that the missing values implied the user had not performed any action, and hence the value was zero. The imputation of missing data with zero resulted in the data being highly sparse and non-linear. This nature of data meant that we had to make two-step tree-based models - one model to predict those with non-zero LTV and the other model to predict the exact value of LTV.

b) Variable reduction and transformation

After eliminating the missing values, the next task was 'feature engineering' (transforming the raw data to improve the performance of machine learning algorithms). We engineered features to understand user behavior immediately after their first login.

c) Model building

The first step was to predict whether the user would generate any revenue in two years. The second step involved categorizing the existing users into three categories, high, medium, and low value.

The data had many high-cardinality categorical variables (which can contain a high number of unique values). We transformed them to a lower-cardinality variable by clubbing similar categories together.

The plan was to build two models in both parts of the model building stage. The first model in both parts would be a decision tree classifier to get actionable insights. The second model would be a tree-based ensemble classifier built to optimize the model's accuracy.

d) Model Comparison & quantifying the impact

Part 1, Model 1: Decision Tree Classification

Within the first part of the modeling, a decision tree that was trained on the in-time data showed that the users not belonging to Segment A (anonymized) with more than one distinct loan – mentioned in the graph below – had a 68% chance of generating revenue. At the same time, users acquired through Segment A (anonymized) who deposited more than $53 had a 72% chance of generating revenue for Kiva.

The directional insights given by the decision tree did not prove optimal when it came to prediction accuracy.

Part 1, Model 2: XGBoost Classification

We trained an XGBoost model on the data to improve the performance of the decision tree classifier. The XGBoost was an improvement to the decision tree classifier and gave a better accuracy score along with a considerable increase in the recall rate in predicting the revenue-generating users.

Part 2, Model 1 - Decision Tree Classification

In the second part of the model, a decision tree was implemented only on the data for the revenue-generating user base. The model revealed that users who deposit more than $Y1 (anonymized) and a subsequent donation of greater than $7.6 have a 72% chance of generating high revenue.

On the other hand, the model also determined that the users who deposit more than $Y2 with a subsequent donation of less than $7.6 had a 67% chance of generating high revenue.

Part 2, Model 2: LGBM Classification

The decision tree built for insights had an accuracy of only 47% that we deemed unacceptable. An LGBM classification model was trained on data to counter the low accuracy score of the decision tree classification. The LGBM model had an accuracy of 75%, a substantial improvement to the decision tree classifier.

Step 5 in BADIR Framework: Actionable Recommendations based on key insights

Recommendation, the final step in the BADIR framework is about making actionable recommendations that would help Kiva create a significant impact and make informed, data-driven decisions.

The models helped us realize that minor changes in user behavior can lead to an increase in the '2-years LTV'. For example, if segment1 (anonymized) lenders are encouraged to make a subsequent loan in the first three months, the users may become high revenue users.

If only 10% of segment1 users could make one more loan, Kiva's '2-year LTV' could potentially rise by $131,000. We also estimated that if Kiva could successfully encourage only 10% of its user base to donate more than $7.5, the '2-year LTV' would increase by $457,000.

The two recommendations together could lead to a potential impact of greater than $500,000 additional revenue for Kiva.

HOW THE BADIR FRAMEWORK HELPED KIVA IDENTIFY $500K+ IN INCREMENTAL REVENUE USING LTV

BADIR, with its step-by-step framework, helped us focus on potential parameters that can drive impact for Kiva while not getting buried in all the available data. Also, it helped us build personas around the users to classify them into various subsets. For example, identify patterns in users who hold the highest probability of subsequent lending. These users will eventually be classified as high-value customers.

Identifying user persona in business settings is akin to knowing every future move of the so-called user. In the case of Kiva, we predicted when a potential lender would lend a subsequent loan in the future and helped them spot areas that needed more effort in terms of profit maximization.

With the data-driven findings, we suggested that minor changes be made to the user flow to benefit from a predicted impact of '2-year higher LTV'. The resultant impact of the BADIR-based solution could be as high as half a million dollars for Kiva.

ABOUT ARYNG

Aryng is a Data Science consulting, training, and advising company. Aryng's SWAT Data science team helps solve complex business problems, develop the company's Data DNA through Data Literacy programs and deliver rapid ROI using machine learning, deep learning, and AI. Our client list includes companies like Google, Box, Here, Applied Materials, Abbott Labs, and GE.

Aryng