Capturing High Value Customers among Migo's Newly Registered Users for Conducting Precision Marketing

Migo x Team 2

January 11, 2020

Date Performed:	December 30, 2019
Group Members:	Wen-Teng Chang 108024507
	Bill Bennett 107065437
	Yu-Ting Lin 107077506
	Tzu-Lun Lin 108073512
Instructor:	Prof. Galit Shmueli

Executive Summary

Migo is an online-to-offline video provider startup focuses in Philippines market. Customers can download video online within Migo platform and watch it offline. Due to data (wifi) is not prevalent in Philippines, Migo has collaborated with several stores (hotspot) to provide the wifi service for their customers to download videos. Migo's strategy now is to expand its market share by providing lots of free trial opportunities to the newly-registered customers. However, how to retain those customers may serve as a critical issue.

Since retaining newly registered users is a critical issue for Migo, not to mention retaining newly registered who tend to create more value to Migo, therefore, we would like to provide Migo a business solution which is about capturing potential high value customers among all newly registered users. To speak precisely, **Migo's marketing team could conduct "less discriminative" precision marketing measure to capture 44% of potential high value customers by marketing to only 20% of total newly registered customers based on our analysis result.** The definition of high value customers referring to those who will spend more than 75% of total newly registered customers' spends collected from September to November, and that equals to 100 pesos.

The response variable is Next_spend_bin referring to a newly registered customer's spending in next month. This is a binary variable which 1 represents those who are high value customers and 0 is the opposite. We have derived several features by ourselves which comes from 3 data sources, transaction, title and engagement datasets given by Migo's data science team. The algorithm utilized includes Logistic Regression, Random Forest and Gradient Boosting. Lift is our performance criterion which aligns to the business goal being conducting precision marketing. Therefore, we compare different algorithms' lifts and finally select Gradient Boosting as our preferable model to provide and the result is presented in the previous paragraph.

The next step of Migo on top of our prediction result may compare between the cost of conducting precision marketing to those potential high value newly registered users and the profits they would possibly return to Migo; after then, we can set a break-even point and find out the decent threshold. Let's say, if the threshold is capturing 40% of potential high value customer by just advertising to 20% of total newly registered customers, as 44% is already above 40%, we could take action. Note that 44% is constrained by the size of the dataset we possessed, this percentage could be even higher if the dataset goes larger as the more data we have the stronger algorithm we may construct.

Business Goal

Our goal is to help Migo capture as many newly registered high-value customers as we can. Besides, it is not just simply capturing high-value customers; it's more like facilitating newly high-value customer retention or stickiness by utilizing "less discriminative" precision marketing measure. It's indeed more useful for a startup company. Due to it aligns with the primary task of Migo which is expanding market share.

Data Mining Goal

Response Variable: Next_spend_bin

Next_spend_bin refers to **a binary outcome** which 1 indicates a newly registered high-value customer spending more than 100 pesos. The threshold spend, 100, is calculated from the upper quartile of all the newly customers' spends during September to November. **We define those who spend more than 100 pesos as high-value customers.**

Predictors and Goal:

The features we use for predicting are extracted from the pre-cleaned transaction, engagement, and title datasets. In addition, we just pick up those who are at their first time transaction_visit. To simply put, our data mining goal is to use self-designed features to predict whether a newly registered Migo's user is a potential high-value customer or not.

Data Description

We derive both training and testing datasets which training set includes 2135 rows and 24 columns and testing set includes 538 rows and 24 columns. Below **Figure 1** is the sample of training dataset and **Figure 2** is the description of response variable and features.

id	Next_Spend	Next_Spend_bin	this_spend	hotspot_id_bin	hotspot_times	product_id_distinct	top_up	transaction_visit_max	paid_transaction_visit_min	paid_transaction_visit_max	is_Firstpay_in_this	Weekday	Weekday_times
10327125484396	25	0	175	1	9	4	275	4	1	4	1	7	5
353143088232096	290	1	881	1	24	3	1631	7	1	5	1	4	25
353729089292046	125	1	105	1	6	4	225	3	1	2	1	5	4
356316022918155	50	0	25	1	1	1	25	1	1	1	1	3	1
356872069917088	25	0	105	1	6	4	225	1	1	1	1	4	6

id	Total_records	content_type_Movie	content_type_Serial	content_type_Episodic	episode_number	running_length	major_migo_genre_bin	migo_genre_bin	content_provider_bin	country_bin	country_group_Local	country_group_Western	country_group_Regional
10327125484396	9	1	0	0	0	107	1	1	1	1	1	0	0
353143088232096	37	1	0	0	0	107	1	1	1	1	0	1	0
353729089292046	6	0	1	0	2	30	1	1	1	1	0	1	0
356316022918155	1	0	1	0	1	30	1	1	1	1	1	0	0
356872069917088	6	0	1	0	1	30	1	1	1	1	1	0	0

Figure 1: Sample Dataset

	Next_Spend	total spend in next month	12	Total_records	how many observations include that unique_id in a month
Response	Next_Spend_bin	binary output; (0,1) = Next_Spend (<=75% , >75%) Sept to Nov Spend	13	content_type_Movie	one hot encoding of content type = Movie
1	this_spend	total spend in this month	**14**	content_type_Serial	one hot encoding of content type = Serial
2	hotspot_id_bin	the most frequent hotspot; binary predictor; (0,1) = A(hotspot) (<=0.18 , >0.18)	**15**	content_type_Episodic	one hot encoding of content type = Episodic
3	hotspot_times	how many times for a certain id transacts at that frequent hotspot	**16**	episode_number	episode number
4	product_id_distinct	how many different product used	**17**	running_length	length of video
5	top_up	price of the product	**18**	major_migo_genre_bin	binary predictor; (0,1) = A(major genre) (<=0.15 , >0.15)
6	transaction_visit_max	total times of transaction in a certain month	**19**	migo_genre_bin	binary predictor; (0,1) = A(genre) (<=0.16 , >0.16)
7	paid_transaction_visit_min	first time payment in a certain month (0 or 1)	**20**	content_provider_bin	binary predictor; (0,1) = A(content provider) (<=0.17 , >0.17)
8	paid_transaction_visit_max	total times of payment in a certain month	**21**	country_bin	binary predictor; (0,1) = A(country) (<=0.16 , >0.16)
9	is_Firstpay_in_this	binary predictor; (0,1) = paid_transaction_visit_max (= 0, > 0)	**22**	country_group_Local	one hot encoding of content type = Local
10	Weekday	the weekday that transacts the most in a certain month	**23**	country_group_Western	one hot encoding of content type = Western
11	Weekday_times	how many times for a certain id transacts on that weekday in a certain month	**24**	country_group_Regional	one hot encoding of content type = Regional
Remark1 : Notation A(predictor) = (# of observations with Next_Spend = 1 under that predictor's category) / (total # of observations under that predictor's category)					
Remark2 : The probability thresholds on response, 2, 18-21, are decided according to EDA which equally separate the categories into 0 and 1.					

Figure 2: Response Variable and Features Description

Data Preparation

Our data preparation can be divided into two parts, which first part will briefly explain how do we construct our training and testing set step by step; in the second part, we'll explain feature engineering techniques we conduct by elaborating the derived features we use for predicting.

Dataset Construction (Appendix 1)

The data we used for analytics comes from three different sources which are transaction, engagement and title datasets. We derived our own features (predictors) based on above datasets to predict the next month spending for newly registered customers.

1. To begin with, we separate the transaction dataset into 4 different months from Sept. to Dec.
2. **Then, we just keep those customers who are first time using Migo platform in a certain month and make sure they do still have transaction in the next month.** In fact, this is the main reason why our size of dataset shrinkages that much.
3. Furthermore, our training dataset is composed of training 1 and training 2, which training 1 has October spend as response variable and September features as predictors. Training 2 is built in the same way but with November response and October features. Therefore, we can combine training 1 & 2 into the final training set since that what we care about is the relative time (this month and next month) but not the actual time (Sept. and Oct.). Testing set has December spendings as response variable and November features as predictors.
4. Finally, we can use unique_id in transaction dataset to map the title_id in engagement dataset, later on, use the title_id to map the features in title dataset. After doing some one-hot encodings (dummy variables) and binary transformations on these newly added features, we succeed constructing our training and testing datasets.

Feature Engineering (Appendix 2)

In this part, we'll select some predictors to elaborate the feature engineering techniques we utilize which will be provided in **Appendix 2**.

Data Mining Solution

We'll divide this section into 3 parts which are Introduction of Performance Evaluation, Models Comparison, and Best Model Interpretation.

Introduction of Performance Evaluation - Lift Analysis (Appendix 3)

We'll carry out lift chart as our measure tool of performance. The basic concept of lift is to compare the # of positive outcome in each decile before and after modeling. The criterion is that the higher the lift of a model, the better it is. Below is the step by step explanation of the lift analysis we are going to apply for models comparison:

1. Predict each observation's propensity of being positive (Next_Spend_bin $= 1$) and sort them in descending order.
2. Equally divide all observations into 10 segments then count total positive outcomes for each segment.
3. Calculate the lift by the formula $\frac{\text{total positive outcome after sorting (modeling)}}{\text{benchmark}}$ for each decile, which benchmark refers to the average of positive outcome in testing data* $(\text{total}/10)$. The reason that average is the decent benchmark is because it is the naive prediction.
4. Compare first two deciles' lifts for different models based on the testing data* and select the model which produces the highest lift as the model we can adopt for future data prediction.

Models Comparison - Logistic Models x Random Forest x Gradient Boosting (Appendix 4)

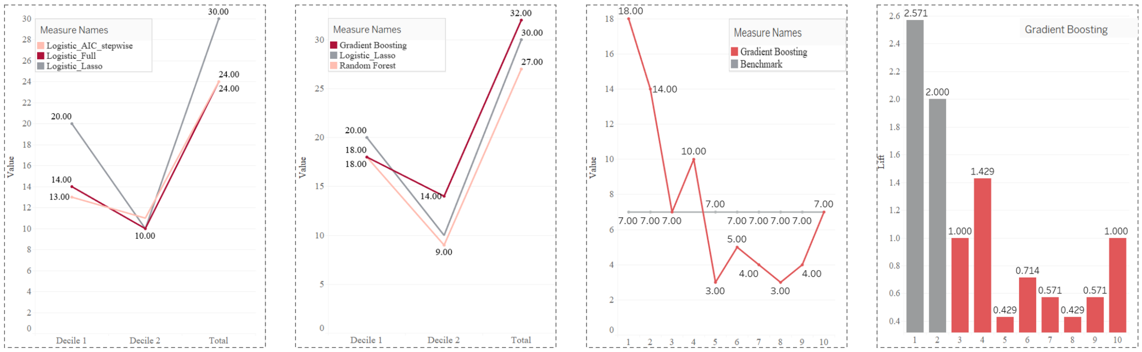

Figure 3: ***Left 1*** *- LRs Comparison /* ***Left 2*** *- Lasso x Rf x Gb /* ***Right 2*** *- Gradient Boosting x Benchmark /* ***Right 1*** *- Gradient Boosting Lifts*

- In the above leftmost figure, we compare the full, stepwise-AIC selected and Lasso selected Logistic Regression models. We discover that Logistic Regression with Lasso shrinkage performs the best, which the summation of first two deciles is 30 outweighs the other two models which just capture 24 newly-registered high value customers. Note that total amounts of newly-registered customer is $(108 = 538 \times \frac{2}{10})$ in the first two decile.
- In the second left of **Figure 3**, we compare Logistic Regression selected by Lasso penalty, Random Forest and Gradient Boosting algorithms, we find out that although Logistic Regression with Lasso shrinkage outperforms Random Forest algorithm, it is worse than Gradient Boosting. Note that if we change our criterion to compare only the first decile lift only, Logistic Regression with Lasso shrinkage will get the throne. By applying Gradient Boosting algorithm, we can capture 32 high value newly-registered customers from the total of 108 customers.
- Both of the right figures present the comparison between gradient boosting (after modelling) and Benchmark (before modelling). The second from the right shows that the amounts captured by gradient boosting is more than those captured by Benchmark $(32 > 14)$. The rightmost figure refers to the lift for each decile, the actual lift for first two deciles is equal to $2.2857 = 32/14$ being decent.

Note: The variables selected for each model are provided in Appendix 4.

Best Model Interpretation - Gradient Boosting Algorithm

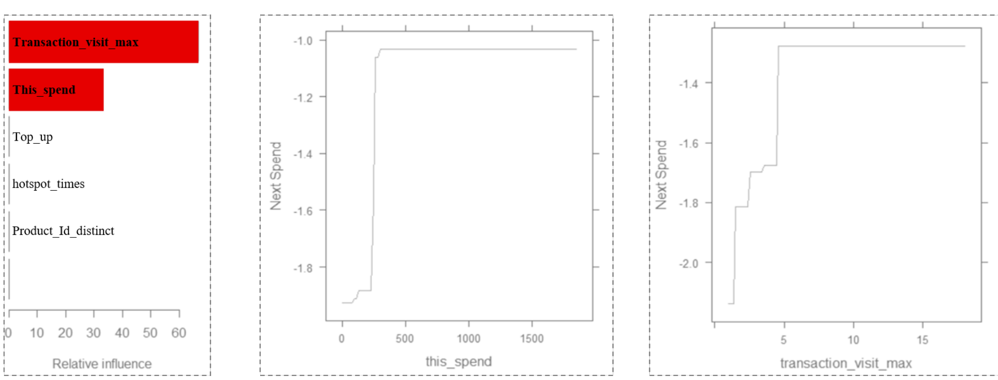

Figure 4: ***Left*** *- Relative Influence /* ***Middle*** *- this_spend Marginal Effect /* ***Right*** *- transaction_visit_max Marginal effect*

- The basic concept of Gradient Boosting is that we first utilize a weak learner on the training data and validate it by validation data, and then we fit the misclassification (or residuals) again and validate it again, by keeping on iterating this process, we can achieve a better learner. Since we first carry out a weak learner, the issue of overfitting doesn't occur easily.

- We can discover the important features from **Figure 4**. From relative influence plot, we unveil that this_spend and transaction_visit_max are critical features; as they increase, the outcome Next_spend tends to increase which are presented in marginal effect plots.

Recommendation - Precision Marketing Measure (Appendix 5)

Our suggestion is to conduct a "less discriminative" precision marketing measure to capture those newly registered high value customer as a short term goal. It could be any, here is one possible marketing solution.

1. We understand that Migo is still in startup stage and may not want to conduct "discriminative" precision marketing which is intuitively considered as a conflict of the short term goal being actively expanding market share. However, we recommend Migo to **imitate "Facebook top fan badge" which is a less drastic means of conducting precision marketing being hardly leading to discrimination**, at least we seldom hear someone feels unfair or even stop using Facebook just for not getting a top fan badge from Facebook fan page. But for those who get, they'll feel themselves more privileged and may somehow enhance their stickiness.
2. **Our recommendation is designing a top fan recognizing system on the Migo platform, let's call it "Migo Badge".** Indeed, this may not derive additional cost due to Migo already has talented software engineers. After completing this recognizing system, Migo could simply grant this "Migo Badge" for those top 20% of customers where includes most high value newly registered customers. In the short run, this privileged badge doesn't need to be materially beneficial to those selected customers. But in the long term, Migo could add value onto it, for example, discount, coupon these kinds of discriminative measure.
3. The last step is to evaluate the feasibility. Migo should estimate the cost of conducting this strategy and the profits expected to gain back from those high value customers. We can help transfer this financial break-even point into a threshold data mining lift; if our algorithm can capture more newly registered high-value customers than the threshold, it's time to get the ball rolling.

Conclusion and Implementation (Appendix 6)

Advantage and Disadvantage-

By applying Gradient Boosting algorithm, Migo can capture 44% of high value customers by conducting precision marketing to 20% of its newly registered customers. Not to mention the algorithm can even be rebuilt with larger data. Thus, if Migo would like to conduct such strategy, it is scientifically proved to be viable. In the short run, we would suggest carrying out a "less discriminative" marketing strategy avoiding people who aren't being marketed feeling unfair.

However, we simply build the model based on training and testing datasets being not enough rigorous. Migo may include validation dataset for a better data mining process. Please find **Appendix 6** for concrete elaboration.

Implementation -

This predicting model can be conducted in the end of each month once the data is well collected, Migo can thereby decide whom to grant "Migo Badge". To grant or not to grant or even how many percentage customers to grant, both of these could be determined by the probabilities predicted, if all the newly registered have low possibilities of being high-value customers, it is ok not to take any actual action although we apply the model technically.

Appendix

Appendix 1 - Dataset Construction

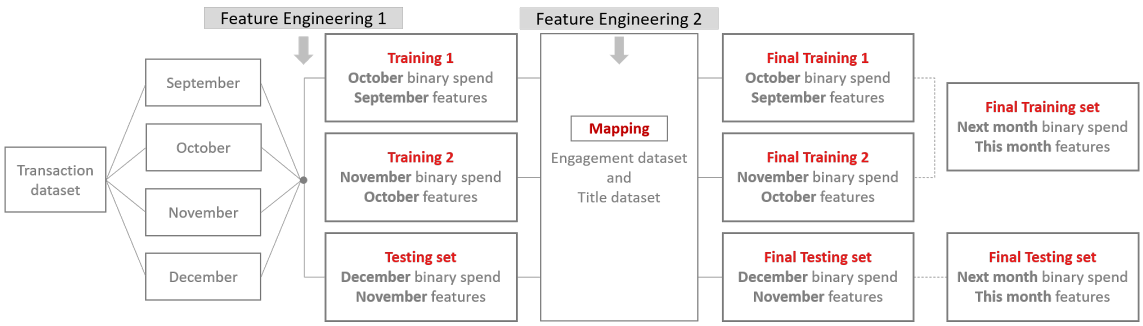

Figure 5: Dataset Construction

Appendix 2 - Feature Engineering

- *Next_spend_bin* : Sum over all the spends grouping by same unique_id, and assign 1 to those customers who spend more than 75^{th} percentile (or 100) of all the customers' spends from September to November, 0 to the rest of customers.
- *hotspot_id_bin and hotspot_id_times* : The first step is to count the number of each hotspot_id that an unique_id transacts at in a certain month and pick up the most frequent hotspot_id, furthermore, the corresponding visiting times to that hotspot_id is our another feature hotspot_times. For the sake of parsimony, we absolutely don't put all 112 distinct hotspot_ids into the model; we bin hotspot_id into binary variable by choosing an optimal threshold, this threshold is determined by the possibility of tendency to predict Next_spend_bin $= 1$.

Appendix 3 - Introduction of Performance Evaluation - Lift Analysis

1. In fact, using testing data to evaluate model performance is not enough rigorous, what we should use is validation data; testing data only serves as unknown data, we can only use it to check the performance result, but never use it for tweaking or adjusting the model.
2. However, in this case, our size of dataset is constrained by Migo, if we holdout a part from training set for validating, we can expect that our performance will get worse. Fortunately, this disadvantage could be offset as the sample size goes large, that is, training, validation, testing set could be applied and the performance will look decent if we have larger dataset. In conclusion, although we just construct training and testing datasets, the result is still persuasive.

Appendix 4 - Models Comparison (Hyperparameters Used)

- Logistic Regression with Lasso: $\lambda = 0.042967$ which is achieved from cross validation.
  - Features left: Paid_transaction_visit_max and hotspot_id_bin
- Random Forest: mtry$= 50$, ntrees$= 500$, importance=TRUE
  - Features used: this_spend
- Gradient Boosting: n.trees$= 1000$, shrinkage $= 0.002$, bag.fraction $= 1$
  - Features used: this_spend, top_up, transaction_visit_max, paid_transaction_visit_min, paid_transaction_visit_max, Weekday and Weekday_times

Appendix 5 - Recommendation (Introduction of Facebook top fan badge)

- Reference website: `https://www.yugatech.com/guides/facebook-top-fan-badge-what-is-it-for/#sthash.RCThGIfr.dpbs`
- Some crucial sentences extracted from above website:
  - How does one get the badge, exactly? **Facebook has no precise, clear guidelines on how to get the badge.** Facebook states that users are eligible for the badge if they are intensively active on a Page.
  - **Having a top fan badge these days on Pages is something that users are proud to show off**, as it displays how interactive they are with the Page. But really, underneath all the " I am a Top Fan! " celebrations, is there a benefit to having the badge? For users, there honestly is not much of an advantage. The badge may be an aid in getting them recognized by fellow users or by the Page, and **mayhaps even give them a level of satisfaction that they have a status others do not have.**
- The similar measure Migo may follow is:
  - Advertise this new recognizing system to the customers without informing them the standard for being granted or not granted "Migo Badge", however, the standard behind is the Gradient Boosting result we provide. We grant all the top 20% newly registered potential high value customers "Migo Badge".
  - Furthermore, in the long run, we could consider creating a hierarchical systems by having different kinds of "Migo Badge" which are bronze, silver and gold or 1 - 5; those who are granted by our algorithm is a bronze one. Further upgrade requires more engagements. This is already far from our topic, just for reference.

Appendix 6 - Conclusion and Implementation

Below figure is the formal process that Migo may follow (assume that we have 3-year data or 30000 observations after datasets construction):

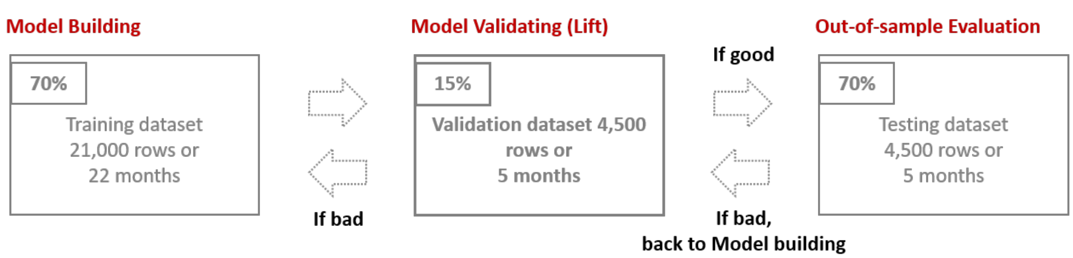

Figure 6: Formal Process of model building - include validation dataset (70-15-15)

Below are some other points which could be further adjusted and improved:

1. The hyperparameters chosen in the Gradient Boosting Model or even the algorithm itself. **(Appendix 4)**
2. The features applied in the models can be based on domain knowledge instead of just driven by data.